AJUR

American Journal of
Undergraduate Research

Volume 20 | Issue 4 | March 2024

www.ajuronline.org

Print Edition ISSN 1536-4585
Online Edition ISSN 2375-8732

AJUR
American Journal of
Undergraduate Research

Volume 20 | Issue 4 | March 2024 | https://doi.org/10.33697/ajur.2024.100

American Journal of Undergraduate Research (AJUR) *is a national, independent, peer-reviewed, open-source, quarterly, multidisciplinary student research journal. Each manuscript of AJUR receives a DOI number. AJUR is archived by the US Library of Congress. AJUR was established in 2002, incorporated as a cheritable not-for-profit organization in 2018. AJUR is indexed internationally by EBSCO and Crossref with ISSNs of 1536-4585 (print) and 2375-8732 (web).*

On Sample Size Needed for Block Bootstrap Confidence Intervals to Have Desired Coverage Rates

Mathew Chandy, Elizabeth D. Schifano, & Jun Yan*

Department of Statistics, University of Connecticut, Storrs, CT

https://doi.org/10.33697/ajur.2024.101

*Student: mathew.chandy@uconn.edu**
Mentors: elizabeth.schifano@uconn.edu, jun.yan@uconn.edu

ABSTRACT
Block bootstrap is widely used in constructing confidence intervals for parameters estimated from stationary time series. Theoretically, the method should provide valid confidence intervals as the length of the time series goes to infinity. In practice, however, it is necessary to know how large of a finite sample is required for block bootstrap confidence intervals to work well. This study aims to answer this question in a simple simulation setting where the data are generated from a first-order autoregressive process. The empirical coverage rates of several commonly used bootstrap confidence intervals for the mean, standard deviation, and the lag-1 autocorrelation coefficient are compared. A quite large sample is found necessary for the intervals to have the right coverage rates even when estimating a simple parameter like the mean. Some block bootstrap methods could fail when estimating the lag-1 autocorrelation. It is surprising that the coverage property even deteriorates as the sample size increases with some commonly used block bootstrap confidence intervals including the percentile intervals and bias-corrected intervals.

KEYWORDS
Autocorrelation; Bias-Correction; Centering; Dependent Data; Percentile; Resampling; Simulation; Time Series

INTRODUCTION

Block bootstrap is a tool to construct confidence intervals (CI) to make inferences about dependent data. Essentially, it depends on correct estimation of the uncertainty in the estimation, similar to the standard bootstrap[1], but for serially dependent data. Early ideas of block bootstrap were developed not long after the standard bootstrap.[2–4] It has since been applied in various fields, for instance, econometrics and meteorology.[5, 6] Block bootstrap is especially useful for serially dependent data when the serial dependence is not specified or not of primary interest. The method is expected to produce CIs with coverage rates matching their nominal levels as the sample size grows.[7] However, when dealing with finite sample sizes, an important question is how large the sample size must be for block bootstrap CIs to have the desired coverage rates.

Lahiri[8] finds that moving block bootstrap has better performance than non-overlapping block bootstrap. Additionally, moving block bootstrap with nonrandom block sizes results in lower mean-squared errors than moving block bootstrap with random block sizes. Buhlmann and Künsch[9] notes that a drawback of block bootstrap is that it heavily depends on block size, which has to be chosen by the user of the method. Even when using the appropriate settings, as noted by Buhlmann[10] observes some general drawbacks of block bootstrap — with respect to how reasonably it imitates the data-generating process. In addition, although block bootstrap is primarily used for stationary time series, it can be outperformed by other bootstrap schemes for linear time series and categorical processes. Still, Buhlmann[10] emphasizes

that a significant advantage of block bootstrap is its simplicity. To be more specific, the resampling step of block bootstrap is not computationally more difficult than the resampling step of basic bootstrap. Furthermore, block bootstrap performs better than local bootstrap in terms of mimicking dependence structures.

For independent data, extensive research has explored the effectiveness of bootstrap standard errors in providing accurate uncertainty measures. For example, Hesterberg [11] observes that while percentile-based CIs for the mean parameter are more accurate than t-intervals for larger sample sizes, their accuracy diminishes for smaller sample sizes. The optimal parameter estimation of a distribution, according to Chernick and Labudde [12], depends on the sample size, the number of bootstrap replicates, and the confidence level. In structural equation modeling, Nevitt and Hancock [13] find that a sample size of 200–1000 is sufficient for interval estimation using standard nonparametric bootstrap. In estimating variance components, Burch [14] reports that as the sample size increases under a normal distribution, nonparametric bootstrap methods approach the coverage of a pivotal quantity, but for other distributions, the coverage can deteriorate. In estimating the correlation coefficient of bivariate normal data, Puth et al. [15] note that even for a sample size of 100 with true correlation coefficient 0, bootstrap methods are less accurate than the Fisher's transformation. The prevailing consensus highlights the necessity of a substantial sample size for bootstrap CIs to attain the desired coverage.

Limited research has offered practical guidance concerning the requisite sample size for employing block bootstrap inference with dependent data. In the context of linear regression involving dependent data, where regression errors stem from a homoscedastic autoregressive process of order-1, the investigation conducted by Goncalves and White [16] reveals that, in cases of small sample sizes, standard error estimates derived from the moving block bootstrap approach may demonstrate greater accuracy than those based on closed-form asymptotic estimates. Nonetheless, even when considering a substantial sample size of 1024, confidence intervals generated through the moving block bootstrap method still fail to adequately encompass the target parameter. The scarcity of existing literature addressing the necessary sample sizes conducive to the efficacy of block bootstrap techniques has spurred the initiation of the present study.

The goal of this paper is to provide recommendations on necessary sample size for block bootstrap with dependent data, similar to what was done for basic bootstrap in Hesterberg [11]. We consider a simple situation of a stationary time series, where the parameters of interests are the mean, standard deviation, and the first-order autocorrelation coefficient. We compare six variants of block bootstrap CIs from the literature: [17, 18] a standard normal CI, a Student's t CI, a percentile CI, a bias-corrected CI, a bias-corrected and accelerated CI, and a recentered percentile CI proposed in this article. Their empirical coverage rates at different sample sizes and dependence levels are compared in a simulation study. The results of this study suggest that recovery of temporal dependence parameters is reliant on the type of interval used.

The remainder of the paper is organized as follows. The first section reviews block bootstrap procedures and how to use block bootstrap estimates to construct CIs; a simple CI obtained by recentering at the original point estimate is proposed for comparison. The second section reports a simulation study comparing the coverage rates of six block bootstrap CIs. A discussion concludes in the final section.

BLOCK BOOTSTRAP CIS

Consider a stationary time series $\{X_t : t = 1, \ldots, n\}$ with length n. Our goal is to construct a CI for a parameter θ in the data generating model of the series. Suppose that $\hat{\theta}_n$ is a point estimator of θ based on the observed series. Bootstrap is a powerful approach to construct CIs. If the observations in the series were independent, a standard nonparametric bootstrap procedure would draw a large number B bootstrap copies of the observed data, and calculate a bootstrap point estimate $\hat{\theta}_n^{(b)}$ for each copy $b = 1, \ldots, B$. The uncertainty of $\hat{\theta}_n$ is then estimated by the empirical uncertainty of the bootstrap point estimates. When serial dependence is present, the bootstrap procedure needs to preserve the serial dependence. Block bootstrap was motivated for this situation.

Block Bootstrap

Block bootstrap preserves the serial dependence in the observed data by partitioning the data into blocks and performing bootstrap on the blocks. In particular, consider block size l and, for convenience, suppose that n is a multiple of l such that there are $k = n/l$ blocks. Each block j is $Y_j = \{X_{(j-1)l+1}, \ldots, X_{(j-1)l+l}\}$, $j = 1, \ldots, k$. Then, we sample k blocks of Y_j's from the set $\{Y_1, \ldots, Y_k\}$ with replacement and concatenate the k sampled blocks in the order they are picked to form a bootstrap sample of the data. The formation of the bootstrap sample ensures that the between-block dependence is weak and that the within-block serial dependence is preserved. Because the blocks here are non-overlapping, this bootstrap approach is known as non-overlapping block bootstrap, or simple block bootstrap.

Alternatively, block-bootstrap can be done with overlapping or moving blocks. Define moving blocks

$$Z_j = \{X_j, \ldots, X_{j+l-1}\}, \quad j = 1, \ldots, n-l+1.$$

Now we draw k blocks from the $(n-l+1)$ blocks of Z_j's with replacement and then align them in the order they were picked to form a block bootstrap sample. If n is not a multiple of l, the last block selected will be reduced in size so that the final size of the block bootstrap sample is n. It is also possible to implement moving block bootstrap while allowing blocks to wrap around the end of the series. In other words, define moving blocks (assuming $l > 1$) as:

$$Z_j = \begin{cases} \{X_j, \ldots, X_{j+l-1}\}, & \text{if } j = 1, \ldots, n-l+1, \\ \{X_j, \ldots, X_n, X_1, \ldots, X_{j-n+l-1}\}, & \text{if } j = n-l+2, \ldots, n. \end{cases}$$

This version does not require that n/l be an integer.

The block size l needs to be chosen with care. It should be large enough for each bootstrap sample to preserve the serial dependence, yet small enough for there to be a large number of blocks to give sufficient variability between each bootstrap sample. As n increases, both l and n/l should also increase. To achieve this, the order of l is often assigned a value as a function of n. A common expression that is considered optimal for the order of l is $\lceil n^{1/3} \rceil$,[9] which was adopted in this study.

Block Bootstrap CIs

Suppose that we have repeated the steps in the last subsection B times, and that for $b \in \{1, \ldots, B\}$, we have obtained a bootstrap point estimate $\hat{\theta}_n^{(b)}$ based on the bth bootstrap sample using the same method that was applied to $\{X_t : t = 1, \ldots, n\}$ to obtain $\hat{\theta}_n$. Now the question is how to construct a CI for θ using the B bootstrap point estimates $\{\hat{\theta}_n^{(1)}, \ldots, \hat{\theta}_n^{(B)}\}$. We consider six kinds of block bootstrap CIs adapted from standard bootstrap CIs.

Standard Normal CI Assuming that $\hat{\theta}_n$ is asymptotically normally distributed with θ as the mean, we just need an estimate of the standard error to construct an approximate CI.[19] Let $\widehat{\text{SE}}$ be the empirical standard error of the bootstrap point estimates $\hat{\theta}_n^{(b)}$ for $b \in \{1, \ldots, B\}$. Let $z_{(\alpha)}$ be the quantile function $F^{-1}(\alpha)$ of the standard normal distribution. A $(1-\alpha)100\%$ standard normal CI is

$$(\hat{\theta}_n - z_{(1-\alpha/2)}\widehat{\text{SE}}, \quad \hat{\theta}_n - z_{(\alpha/2)}\widehat{\text{SE}}).$$

This CI is centered by the original point estimate $\hat{\theta}_n$ and is symmetric. The standard CI is classified by Efron and Tibshirani[19] as a confidence interval based on bootstrap "tables", which essentially means it is based on an asymptotic distribution with an estimated asymptotic variance (standard error). Its validity relies on whether the distribution of $\hat{\theta}_n$ is reasonably well approximated by its asymptotic normal distribution and whether the bootstrap $\widehat{\text{SE}}$ approximates the true standard error.

Student's t CI The procedure for constructing a Student's t CI based on standard bootstrap is described in Efron and Tibshirani[19]. Let $t_{(\alpha,k)}$ be the quantile function $F^{-1}(\alpha, k)$ of a t distribution with k degrees of freedom. With block bootstrapping, a $(1-\alpha)100\%$ Student's t CI is

$$(\hat{\theta}_n - t_{(1-\alpha/2),k-1}\widehat{\mathrm{SE}}, \quad \hat{\theta}_n - t_{(\alpha/2),k-1}\widehat{\mathrm{SE}}),$$

where k is the number of blocks. This CI is centered by the original point estimate $\hat{\theta}_n$ and is symmetric. Like the standard normal interval, the Student's t CI is classified by Efron and Tibshirani[19] as a confidence interval based on bootstrap "tables". In this case, its validity relies on whether the distribution of $\hat{\theta}_n$ is reasonably well approximated by the t_{k-1} distribution with an expected value of θ and whether the bootstrap $\widehat{\mathrm{SE}}$ approximates the true standard error.

Percentile CI The percentile CI was first suggested in Efron[1]. Let $\hat{\theta}^B_{n,\alpha}$ be the empirical 100αth percentile of $\{\hat{\theta}^{(1)}_n, \ldots, \hat{\theta}^{(}_n$ A $(1-\alpha)100\%$ empirical percentile CI is

$$(\hat{\theta}^B_{n,\alpha/2}, \quad \hat{\theta}^B_{n,1-\alpha/2}).$$

This CI is not necessarily centered by the original point estimate $\hat{\theta}_n$. As will be shown in our simulation study, this approach works well for the marginal mean and standard deviation of a serially dependent process, but its coverage of the temporal dependence deteriorates as n increases, which is contrary to what one would expect.

Bias-Corrected (BC) CI The procedure for constructing a bias-corrected Bootstrap CI based on standard bootstrap is described in Carpenter and Bithell[20]. Let $\hat{z}_0 = \Phi^{-1}\{\#\{\hat{\theta}^{(b)}_n < \hat{\theta}_n\}/B\}$ for $b \in \{1, \ldots, B\}$. Define $\alpha_1 = \Phi(2\hat{z}_0 - z_{1-\alpha/2})$ and $\alpha_2 = \Phi(2\hat{z}_0 - z_{\alpha/2})$. A $(1-\alpha)100\%$ BC CI is

$$(\hat{\theta}^B_{n,\alpha_1}, \quad \hat{\theta}^B_{n,\alpha_2}).$$

Bias-Corrected and Accelerated (BCA) CI The BCA CI was first suggested in Efron[21]. Let $Z_{(i)}$ be the original sample without the ith block z_i for $i \in \{1, \ldots, k\}$, let $\hat{\theta}_{(i)}$ be the statistic of $Z_{(i)}$, and let $\hat{\theta}_{(.)} = k^{-1}\sum_{i=1}^k \hat{\theta}_{(i)}$. Let

$$\hat{a} = \frac{\sum_{i=1}^k (\hat{\theta}_{(.)} - \hat{\theta}_{(i)})^3}{6\{\sum_{i=1}^k (\hat{\theta}_{(.)} - \hat{\theta}_{(i)})^2\}^{3/2}}.$$

Define

$$\alpha_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{\alpha/2}}{1 - \hat{a}(\hat{z}_0 + z_{\alpha/2})}\right)$$

and

$$\alpha_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{1-\alpha/2}}{1 - \hat{a}(\hat{z}_0 + z_{1-\alpha/2})}\right).$$

A $(1-\alpha)100\%$ BCA CI is

$$(\hat{\theta}^B_{n,\alpha_1}, \quad \hat{\theta}^B_{n,\alpha_2}).$$

This CI is not necessarily centered by $\hat{\theta}_n$. The BCA method corrects for bias and skewness of the B bootstrap point estimates $\{\hat{\theta}^{(1)}_n, \ldots, \hat{\theta}^{(B)}_n\}$ by including bias-correction and acceleration factors. The acceleration factor refers to the rate of change of the standard error of $\hat{\theta}_n$ with respect to θ.

Recentered Percentile CI We propose a CI that is centered at the original point estimate and uses the variation in the bootstrap estimates to construct the error bound. The motivation behind proposing such an interval was based on the simulation performance of the BC and BCA intervals, which will be discussed further in the Results section. This interval requires the computation of $\bar{\theta}_n^B = n^{-1}\sum_{b=1}^{B}\hat{\theta}_n^{(b)}$, the mean of all bootstrap point estimates. A $(1-\alpha)100\%$ CI is centered around $\hat{\theta}_n$ and can be written as

$$(\hat{\theta}_n + \hat{\theta}^B_{n,\alpha/2} - \bar{\theta}^B_n, \quad \hat{\theta}_n + \hat{\theta}^B_{n,1-\alpha/2} - \bar{\theta}^B_n).$$

It is not necessarily symmetric, as different critical values are used to compute the lower and upper bounds. It has the same width as the percentile CI.

SIMULATION DESIGN

We compared the performance of the different block bootstrap CI methods under two marginal distributions: standard normal and unit exponential.

Marginal Standard Normal Distribution

We generated time series X_t from a 1st order autoregressive (AR(1)) process:

$$X_t = \phi X_{t-1} + \epsilon_t,$$

where ϕ is an autoregressive coefficient, and ϵ_t is a series of independent errors from a normal distribution with mean zero and variance σ^2_ϵ. The strength of the serial dependence is controlled by ϕ, which was set to five levels: $\{-0.4, -0.2, 0.0, 0.2, 0.4\}$. We only used serial dependences as strong as 0.4, because we only seek to establish the general trend as the strength of the autocorrelation increases, and how it varies depending on the sign of the autocorrelation and the parameter of interest. The series X_t has mean zero and variance $\sigma^2_x = \sigma^2_\epsilon/(1-\phi^2)$, so for each value of ϕ, we set $\sigma^2_\epsilon = (1-\phi^2)$ such that $\sigma^2_x = 1$.

Three target parameters of X_t were considered: 1) $\mu = 0$, the mean of X_t; 2) $\sigma_x = 1$, the standard deviation of X_t; and 3) ϕ, the lag-1 autocorrelation coefficient. To investigate the effect of sample size n, we considered an array of values $n \in \{100, 200, 400, 800, 1600, 3200\}$. In each configuration, we generated 10,000 replicates. The block bootstrap sampling step was done with function `tsboot` from R[22] package *boot*,[23] with block size $\lceil n/l \rceil$. This function by default is an implementation of moving block bootstrap as described in the previous section, meaning that that blocks are allowed to wrap around, and we tried both $l = \lceil n^{1/3} \rceil$ and $l = \lceil 2n^{1/3} \rceil$, keeping the order of the block size constant but varying the coefficient. For each replicate, we constructed six 95% block bootstrap CIs for each parameter as described in the last section with $B = 1000$. We can estimate μ, σ_x, and ϕ by computing the sample mean, sample standard deviation, and lag-1 autocorrelation, respectively, of each bootstrap sample. Then we can construct intervals for each parameter using the appropriate procedures described in *Block Bootstrap CIs*. Then we estimated their actual coverage rates along with their 95% confidence intervals from the 10,000 replicates.

The coverage rates of the CIs were used to compare the performance of CIs. Let $\hat{\theta}_{L,r}$ and $\hat{\theta}_{U,r}$ be the lower and upper bound, respectively, for the confidence interval constructed for each replicate $r \in \{1, \ldots, R\}$, where R is the number of replicates. Then the empirical coverage rate is $\sum_{r=1}^{R} I\{\hat{\theta}_{L,r} < \theta < \hat{\theta}_{U,r}\}/R$, where $I(\cdot)$ in the indicator function. If a CI method is valid, then the coverage rate is expected to match the nominal level. Because it is unlikely for the coverage to exactly match the nominal level, we can construct a 95% Clopper-Pearson exact CI of the coverage rate,[24] which is an estimate of a proportion with $R = 10,000$. We used the R *PropCIs* package to achieve this.[25] The choice of Clopper-Pearson was motivated by the Wald interval's poor coverage as the proportion approaches 0 or 1,[26] although when we tried Wald intervals, the coverage rate intervals did not appear to have large differences. If the proportion 0.95 is included in the interval, the block bootstrap method is likely performing well. If all values in the interval are below 0.95, the results would suggest that the method either is providing inaccurate estimation, is underestimating the process' variability, or a combination of both. If all values in the interval are above .95, the results suggest that the method is

overestimating the process' variability. **Figure 1** summarizes the empirical coverage rates and the 95% confidence intervals of the real coverage for a marginal standard normal distribution using block bootstrap with $l = \lceil n^{1/3} \rceil$, generated using the R *ggplot2* package.[27]

Marginal Unit Exponential Distribution

Additionally, to investigate if the results are robust to nonnormal marginal distributions, we evaluated the performance of block bootstrap for a time series with a non-normal marginal distribution. Specifically, we estimated the mean, standard deviation, and the lag-1 autocorrelation coefficient of a stationary series with marginal unit exponential distribution. Note that we expect the CIs that are based on bootstrap "tables" to depend on the asymptotic distribution of the estimator. This asymptotic distribution depends more on the sample size than on the marginal distribution of the time series. So we expect such CIs to have similar performance for different marginal distributions when the sample sizes is large. The percentile-based CIs (Percentile, BC, BCA, Recentered Percentile) are not necessarily expected to perform better under non-normal marginal distributions. The student's t and and normal-based CIs are only noticeably different when the number of blocks is smaller than 20.

The series were generated by marginally transforming the AR(1) series X_t in the first simulation study by

$$W_t = F^{-1}[\Phi(X_t)],$$

where $F^{-1}(p)$ is the quantile function for the unit exponential distribution. The true mean (μ) and standard deviation (σ_w) parameters of W_t are 1. The lag-1 autocorrelation coefficient (ρ) is not invariant to the transformation,[28] but its value can be obtained by

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F^{-1}[\Phi(x)]F^{-1}[\Phi(y)]g_2(x,y;\phi)\mathrm{d}x\mathrm{d}y - 1,$$

where $g_2(x, y; \phi)$ is the density of a standard bivariate normal distribution with correlation parameter ϕ. We kept the configuration of $\phi \in \{-0.4, -0.2, 0.0, 0.2, 0.4\}$, and the corresponding lag-1 autocorrelation coefficients are $\rho \in \{-0.298, -0.156, 0, 0.170, 0.355\}$.

SIMULATION RESULTS

Marginal Standard Normal Distribution

For estimating the mean parameter μ, the top panel of **Figure 1** suggests that all methods eventually approach correct coverage of μ as sample size increases. Student's t CIs appear to need the smallest sample size to achieve correct coverage, except for samples with strong negative dependence, in which case, they actually over-cover μ for smaller sample sizes. For instance, for a sample with $n = 100$ and $\phi = -0.4$, the lower bound for a Students t CI's coverage of μ is greater than 95%, whereas the coverage intervals for other methods contain 95%. The standard normal, percentile, BC, and BCA, and recentered percentile CIs require similar sample sizes to recover μ at the nominal level for all combinations of n and ϕ. All methods seem to require a smaller sample to recover μ at the nominal rate when dealing with negative dependence versus positive dependence. For example, BC CIs recover μ for $n \geq 100$ when $\phi = -0.2$, but they only recover μ for $n \geq 800$ when $\phi = 0.2$. In addition, as a negative dependence gets stronger, holding everything else equal, coverage increases, which lead to the Student t CI's aforementioned over-coverage. As a positive dependence gets stronger, holding everything else equal, coverage decreases, and a larger sample is necessary to recover μ. A possible explanation for this is that if a stationary series has a positive autocorrelation, the effective sample size is decreased, whereas if a series has a negative autocorrelation, the effective sample size is increased.[29] Additionally, this seems to have a greater effect on the the estimation of the location parameter versus that of the scale parameter or temporal dependence parameter.

For estimating the standard deviation parameter σ_x, **Figure 1** suggests that every method can reach nominal coverage of σ_x if the sample is large enough, but for a given n and ϕ, coverage of σ_x will be lower than coverage of μ in general. Like μ, σ_x can be covered by Student t CIs with smaller sample sizes when compared to other methods. Unlike

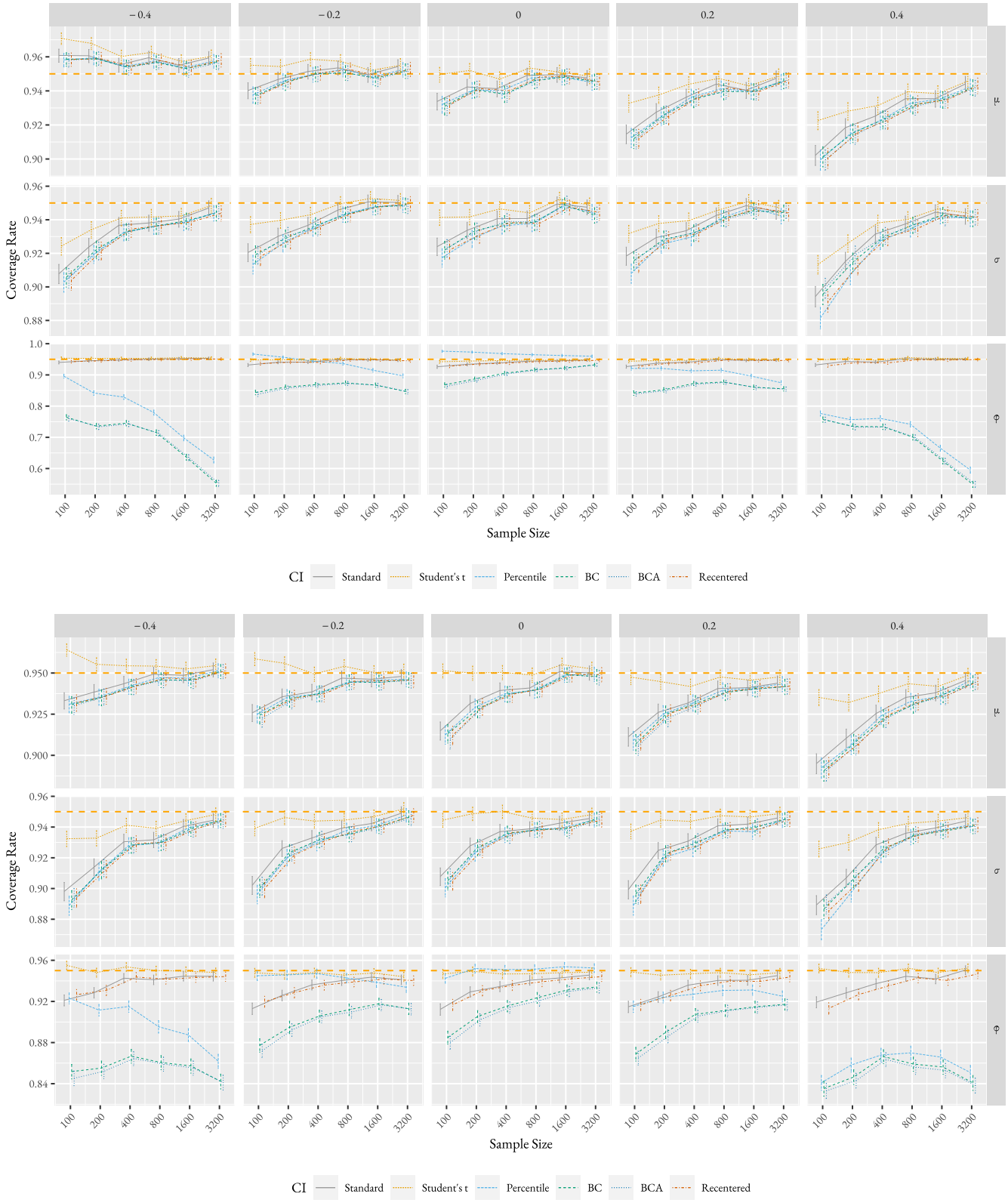

Figure 1. Empirical coverage rates of different 95% block bootstrap CIs for the marginal mean μ, the marginal standard deviation σ_x, and the first-order autocorrelation coefficient ϕ of an AR(1) process with a marginal standard normal distribution with AR coefficient $\phi \in \{-0.4, 0.2, 0, 0.2, 0.4\}$ and series length $n \in \{100, 200, 400, 800, 1600, 3200\}$ based on 10,000 replicates of block bootstrap with $l = \lceil n^{1/3} \rceil$. The error bars represent 95% CIs of the real coverage rates. Top: $l = \lceil n^{1/3} \rceil$. Bottom: $l = \lceil 2n^{1/3} \rceil$.

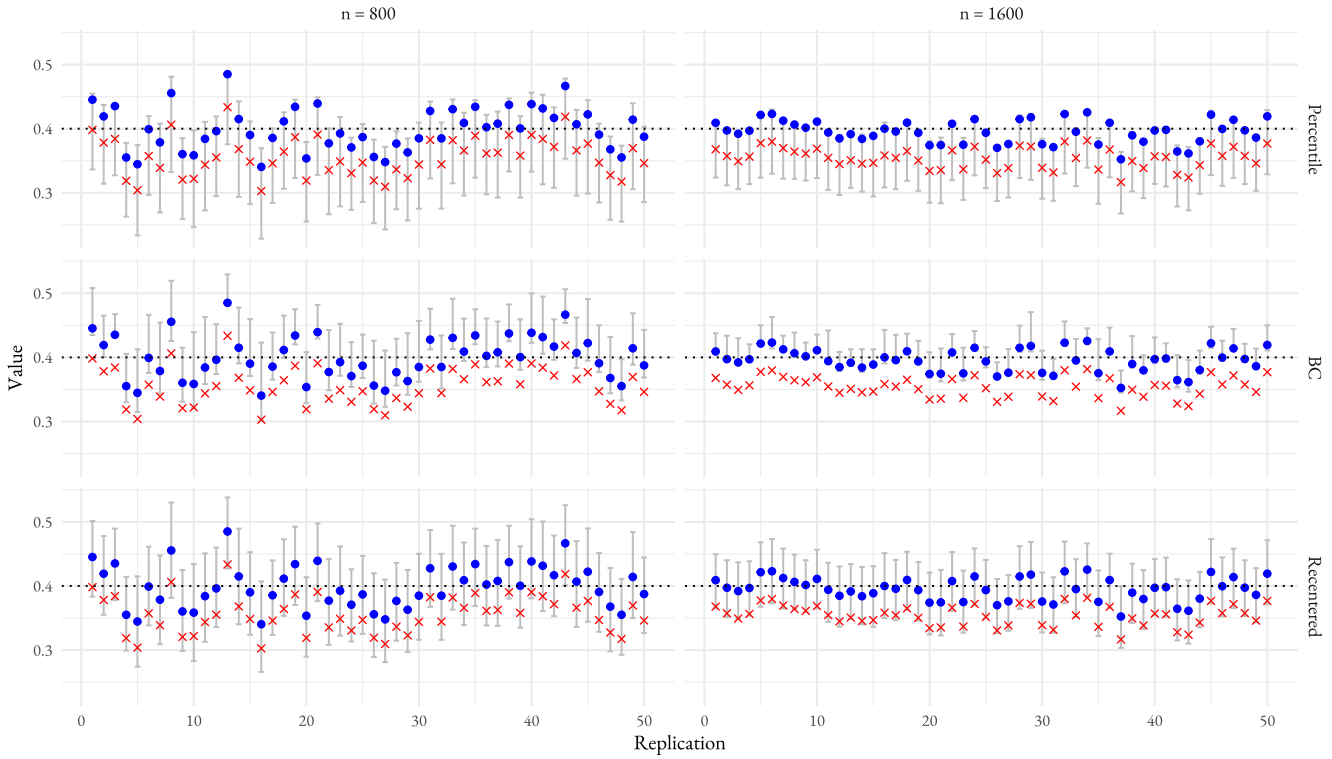

Figure 2. 50 replicate percentile, BC, and recentered percentile CIs for samples of size $n \in \{800, 1600\}$ ($l = \lceil n^{1/3} \rceil$) for the lag-1 autocorrelation of an AR(1) process with $\phi = 0.4$. For each replicate, the lower and upper bounds of the CIs are displayed, as well as $\hat{\theta}_n$ (blue circle) and $\bar{\theta}_n^B$ (red cross).

μ, there is no over-coverage issue for σ_x when $\phi = -0.4$. Standard normal, percentile, BC, BCA, and recentered percentile CIs again have similar performance. All methods seem to have slightly higher coverage of σ_x when ϕ is negative versus when ϕ is positive. Regardless of the sign, coverage of σ_x gets worse as the strength of the temporal dependence increases.

For estimating the autocorrelation parameter ϕ, **Figure 1** suggests that while standard normal, Student's t, and recentered percentile CIs do approach correct coverage as sample size increases, percentile, BC, and BCA CIs deteriorate as sample size increases, especially as the strength of the temporal dependence increases. Because of this, only standard normal, Student's t, and recentered percentile CIs should be considered as effective block bootstrap methods to estimate ϕ. Student's t CIs once again can achieve correct coverage with smaller sample sizes when compared to standard and recentered percentile CIs which perform similarly. Student's t CIs can can recover ϕ at the nominal level for $n \geq 100$ when the sample's temporal dependence is as strong as 0.4. Coverage appears to be higher for all methods when the dependence is negative rather than positive. Whether or not the dependence is negative or positive, coverage of ϕ seems to increase slightly as the absolute value increases for standard normal, Student's t, and recentered percentile CIs. For the values of ϕ observed, there are no examples of over-coverage for standard normal, Student's t, BC, BCA, or recentered percentile CIs. However, percentile CIs appear to over-cover ϕ for smaller sample sizes when $\phi = -0.2$ and when $\phi = 0$, indicating again that they should not be used.

The outcomes of the ϕ estimation raise a natural question about the lackluster performance of certain methodologies. To delve into this inquiry, a set of 50 CIs was generated for each of the percentile, BC, and recentered percentile approaches for samples of $n \in \{800, 1600\}$. Illustrated in **Figure 2**, it becomes evident that the percentile-based CIs exhibit a notable bias, predominantly manifesting as a substantial underestimation of ϕ with point estimator $\bar{\theta}_n^B$, that is, the average of B bootstrap point estimates. As the sample size increases from 800 to 1600, this bias does not vanish while the uncertainty reduces, which explains why the coverage rates deteriorate. The bias in $\bar{\theta}_n^B$ also appears to inval-

idate the bias-correction in the BC bootstrap, leading to the poor performance of the BC intervals. The BCA intervals have the same problem as the BC intervals in the bias-correction step. The root of the issue appears to be that the autocorrelation in the block bootstrap samples is somehow smaller compared to that in the original sample. On the other hand, the original point estimator $\hat{\theta}_n$ is asymptotically unbiased. Since the width is based on the uncertainty in the bootstrap point estimates $\hat{\theta}_n^{(b)}$, $b = 1, \ldots, B$, the percentile CIs recentered at the original point estimate $\hat{\theta}_n$ provide desired coverage.

To summarize, the performance of the CIs depends on the target parameter. When estimating μ and σ_x, any CI will do, although Student's t CIs perform noticeably better than the others. However, when estimating ϕ, the choice of method is of utmost importance as to avoid coverage deterioration. Coverage rates are acceptable at smaller sample sizes when phi is positive versus when phi is negative. In other words, a larger sample size is generally required to estimate a parameter for a sample with a negative ϕ versus a positive ϕ of the same magnitude. In order to know if coverage will increase as the strength of the temporal dependence increases, one need to know what the parameter of interest is, and in the case of μ, the direction of the serial dependence. The BC approach does not seem to be correcting bias appropriately when estimating ϕ. Like the percentile method, the recentered percentile method uses the spread from the bootstrap to construct the width of the CI. However, the recentered approach, does not correct from the original point estimate $\hat{\theta}_n$.

The results for $l = \lceil 2n^{1/3} \rceil$ are reported in the bottom panel of Figure 1. The performances generally seems to be inferior compared those with $l = \lceil n^{1/3} \rceil$, but importantly, the patterns in performance when varying other parameters appear to be robust to the different block size. For negative autocorrelations, the coverage rates of μ appear to be lower when using $l = \lceil 2n^{1/3} \rceil$. For example, whereas $n = 100$ or 200 would seem sufficient for most CIs when using $l = \lceil n^{1/3} \rceil$, $n = 800$ or 1600 is necessary to capture negative autocorrelations for $l = \lceil n^{1/3} \rceil$. Student's t CIs do not seem to be as affected by this change in l: for $\phi = -0.4$ and -0.2, they still over-cover μ for smaller values of n. The results for σ_x with $l = \lceil 2n^{1/3} \rceil$ look very similar to those the results for σ_x with $l = \lceil n^{1/3} \rceil$, but coverage rates of σ_x do look slightly lower especially for negative values of ϕ, although Student's t CIs are again not as influenced by this change in l. A larger sample size seems necessary when using other CIs to estimate σ_x for $l = \lceil 2n^{1/3} \rceil$. Recentered percentile and standard CIs have slightly lower coverage rates when estimating negative values of ϕ with $l = \lceil 2n^{1/3} \rceil$. Although it is still a problem, the coverage deterioration appears to be less dramatic for BCA, BC, and percentile CIs. Aside from these differences, the overall changes in performance when other experimental factors are changed are the same as when $l = \lceil n^{1/3} \rceil$.

Marginal Unit Exponential Distribution

For the scenario of marginal exponential distribution, the empirical coverage rates for μ, σ_w, and the lag-1 autocorrelation coefficient ρ using block bootstrap with $l \in \{\lceil n^{1/3} \rceil, \lceil 2n^{1/3} \rceil$, as well as 95% confidence intervals of the real coverage are displayed in **Figure 3**. Additionally, a set of 50 CIs are displayed for each of the percentile, BC, and recentered percentile approaches for exponentially distributed samples of $n \in \{800, 1600\}$ with lag-1 autocorrelation coefficient 0.355 ($\phi = 0.4$) in **Figure 4**.

It appears that a greater sample size is generally required for the bootstrap CIs to cover the mean and standard deviation parameters in the exponential margin case than in the normal margin case. However, the other trends and patterns discussed regarding the performance of various methods and diverse parameters remain unchanged. For example, Student's t confidence intervals still exhibit higher coverage rates in comparison to alternative methods. Performance continues to be more favorable when temporal dependence is negative rather than positive. Again, altering the block size results in the same changes in performance of different CIs as those in the scenario of marginal normal distribution. Of particular significance, the percentile, BC, and BCA confidence intervals still display a decline in coverage accuracy for the lag-1 autocorrelation coefficient as sample size increases as demonstrated in Figure 4. Both the percentile and BC intervals persist in manifesting the same bias issue. On the other hand, the recentered percentile confidence interval

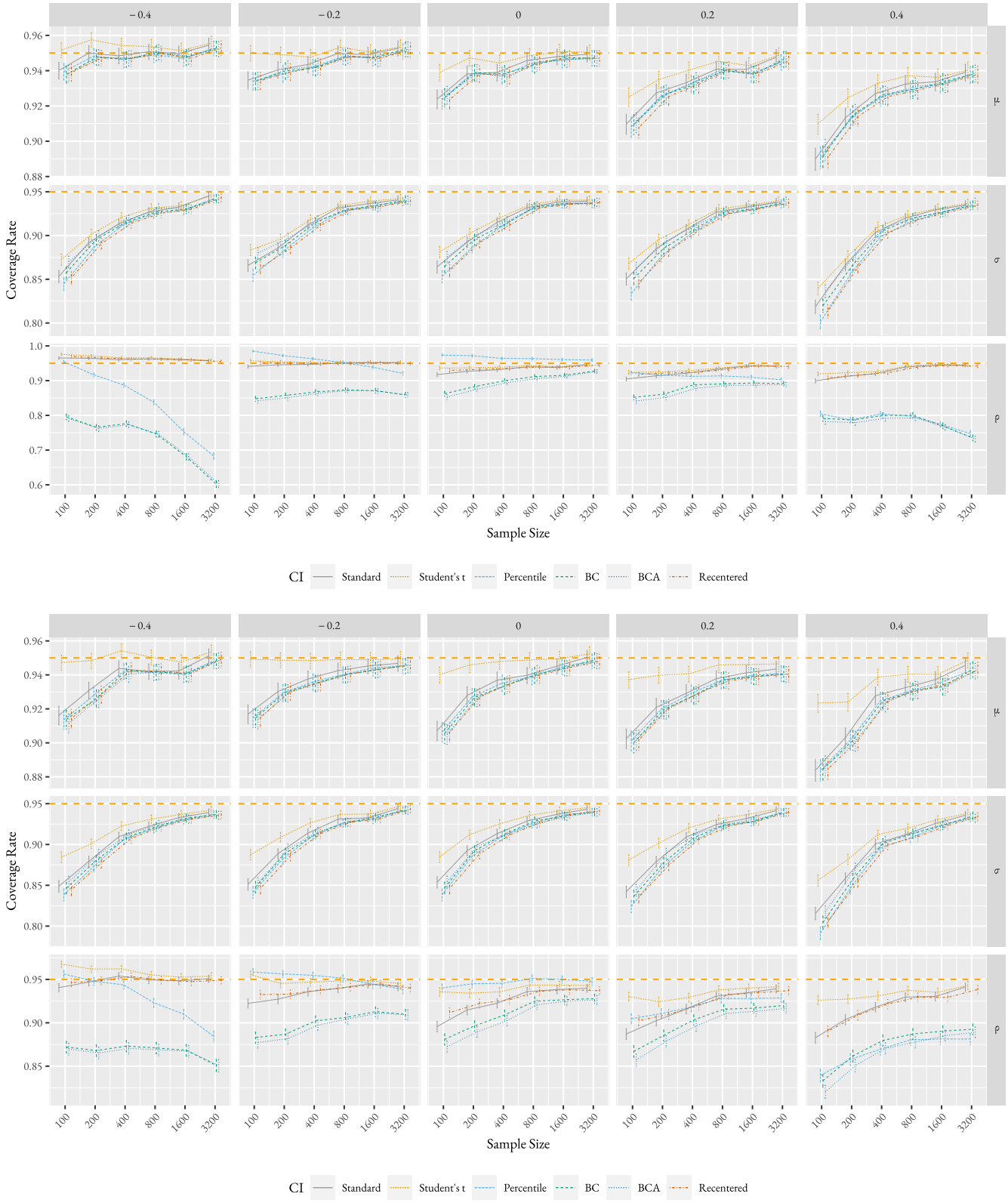

Figure 3. Empirical coverage rates of different 95% block bootstrap CIs for the marginal mean μ, the marginal standard deviation σ_w, and the first-order autocorrelation coefficient ρ of a stationary series with marginal unit exponential distribution obtained by transforming an AR(1) process with $\phi \in \{-0.4, 0.2, 0, 0.2, 0.4\}$ with series length $n \in \{100, 200, 400, 800, 1600, 3200\}$ based on 10,000 replicates replicates of block bootstrap with $l = \lceil n^{1/3} \rceil$. The error bars represent 95% CIs of the real coverage rates. Top: $l = \lceil n^{1/3} \rceil$. Bottom: $l = \lceil 2n^{1/3} \rceil$.

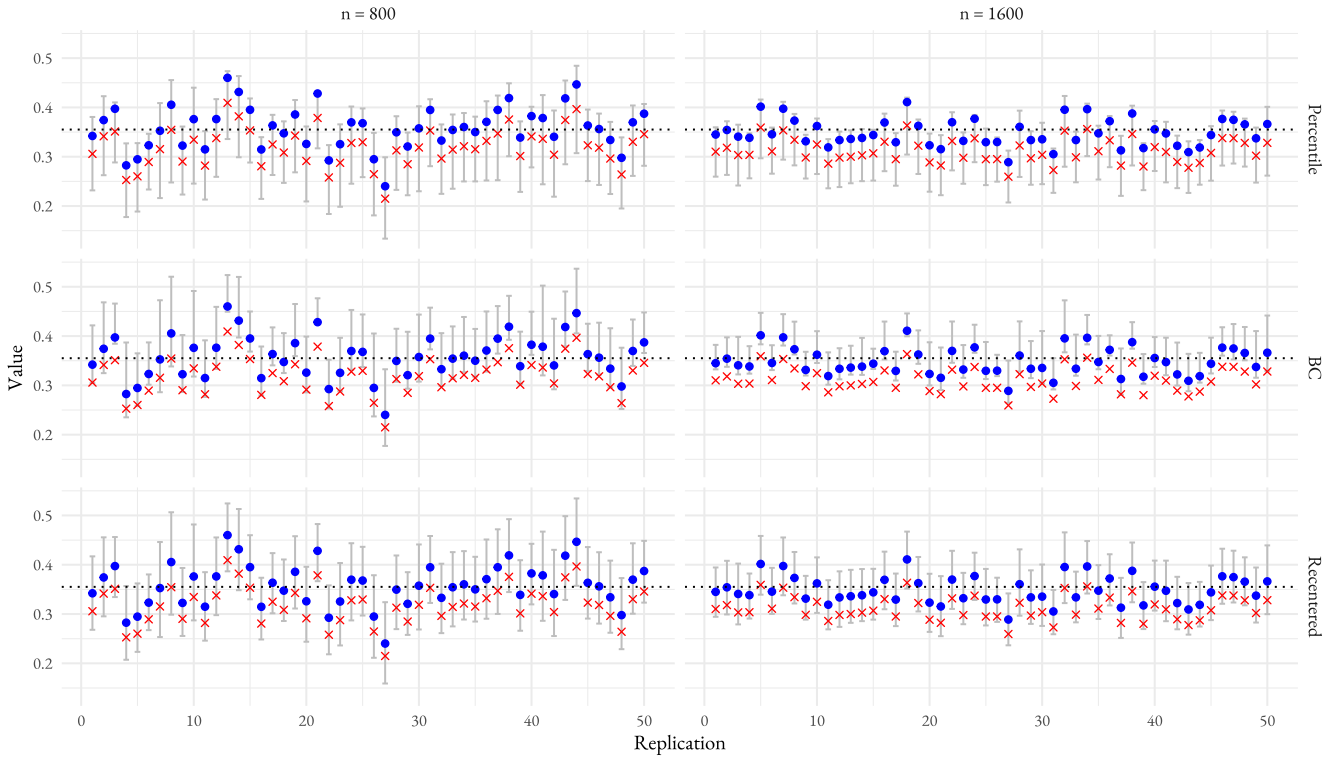

Figure 4. 50 replicate percentile, BC, and recentered percentile CIs for the lag-1 autocorrelation coefficient $\rho = 0.355$ ($\phi = 0.4$) of a stationary series with marginal exponential distribution obtained by transforming an AR(1) process with $\phi = 0.4$ with sample size $n \in \{800, 1600\}$ ($l = \lceil n^{1/3} \rceil$). For each replicate, the lower and upper bounds of the CIs are displayed, as well as $\hat{\theta}_n$ (blue circle) and $\bar{\theta}_n^B$ (red cross).

continues to be effective in estimating the temporal dependence due to the inherent unbiasedness of the original point estimator. In sum, for the most part, the findings for series that are marginally exponentially distributed closely mirror those attained for series that are marginally normally distributed.

DISCUSSION

Block bootstrap is a useful method for estimating parameters of a time series, from simple parameters like the mean to more complicated temporal dependence factors. We know theoretically that the block bootstrap procedure will cover a parameter of a time series at the nominal level given an infinitely large sample,[7] so the goal for this study was to find the smallest finite sample length n of a time series in order for the block bootstrap procedure to recover its associated parameters at an acceptable rate. Our analysis relies on the assumption that there is a size n large enough for the method to work: that is, the method's performance improves as n increases. Out of the six types of intervals used in this study, this assumption was found to hold true with respect to estimating ϕ only for standard normal, Student's t, and recentered percentile CIs, whereas percentile, BC, and BCA intervals exhibited coverage deterioration as n increased. The percentile CI's coverage deterioration can be attributed to bias that is not corrected as n increases. Specifically, as n increases, the width of the CI decreases, but because the percentile CI underestimates ϕ, the coverage decreases. The BC CI seems to correct the bias, but the width of the CI seems to be too short. The acceleration factor of the BCA CI seems to fail, as the width of the CI seems to be too short.

One of the goals of this study was to provide some practical recommendations for necessary sample sizes when using block bootstrap to estimate the parameters of serially dependent data. When using Student's t intervals and the marginal distribution and temporal dependence is unknown, the results of this study suggest that $n \geq 1600$ may be necessary for common practice to estimate μ, whereas $n > 3200$ may be necessary to estimate the standard deviation. Student's t is always preferable to Standard Normal CIs as they performs better for smaller sample sizes and performs

as good or better for larger sample sizes. Lastly, to estimate lag-1 autocorrelation, $n \geq 100$ using the Student's t method may be sufficient under a marginal standard normal distribution, whereas $n \geq 1600$ may be required under under a marginal exponential distribution. Further investigation may be necessary to see if there are other percentile-based interval corrections that fix the coverage deterioration problem for ϕ.

Although we have only used serial dependences as strong as 0.4, we have established the trends as $|\phi|$ gets larger. When estimating μ, we expect coverage rates to decrease as ϕ approaches 1 — as ϕ approaches -1, we may observe increased over-coverage. When estimating the standard deviation, we expect a larger sample size to be necessary as $|\phi|$ gets closer to 1. Lastly, when estimating ϕ for a marginal normal distribution, we expect a larger sample size to be necessary as $|\phi|$ approaches 0, assuming standard normal, Student's t, or recentered percentile CIs are used. However, when estimating the first-order autocorrelation of a marginal exponential distribution using the same methods, we expect coverage rates to respond to stronger dependences in a trend similar to that of coverage rates of μ. We expect other percentile-based CIs, which are already inadequate for relatively weak dependence structures, to perform even worse as $|\phi|$ approaches 1.

This study could be used as a guide for applied statistics courses for students to generally understand how large of a sample size is sufficient for block bootstrap to be used versus other inference methods. For undergraduate or graduate students, block bootstrap is not typically a part of curriculum, but the results of this study can easily be used to demonstrate when it is practical to use this method. This information could also prove to be useful for research using block bootstrap estimation of time series in domains such as econometrics. Future studies could investigate the n needed to make inferences about other forms of serially dependent data such as a moving average process. One could also investigate if there are types of block bootstrap interval construction such as ABC or bootstrap-t intervals[19] that could more appropriately recover the parameters of a time series. We discussed some drawbacks of block bootstrap in the introduction, which could motivate a similar simulation study for alternatives to block bootstrap, such as AR-Sieve bootstrap,[30] which Buhlmann[10] finds to be the best for linear time series. Finally, there is a need for a more in-depth exploration to comprehend the reasons behind the subpar performance of existing percentile-based CIs when estimating the autocorrelation parameter. It is crucial to conduct a thorough investigation into the specific scenarios where the proposed CI demonstrates superior performance and the conditions under which it should be recommended.

REFERENCES

1. Efron, B. (1979) Bootstrap methods: Another look at the jackknife. *The Annals of Statistics* 7(1), 1–26. *https://doi.org/10.1007/978-1-4612-4380-9_41*
2. Hall, P. (1985) Resampling a coverage pattern. *Stochastic Processes and their Applications* 20(2), 231–246. *https://doi.org/10.1016/0304-4149(85)90212-1*
3. Carlstein, E. (1986) The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *The Annals of Statistics* 14(3), 1171–1179. *https://doi.org/10.1214/aos/1176350057*
4. Kunsch, H. R. (1989) The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics* 17(3), 1217–1241. *https://doi.org/10.1214/aos/1176347265*
5. MacKinnon, J. G. (2006) Bootstrap methods in econometrics. *The Economic Record* 82(1), 2–18. *https://doi.org/10.1111/j.1475-4932.2006.00328.x*
6. Varga, L., and Zempléni, A. (2017) Generalised block bootstrap and its use in meteorology. *Advances in Statistical Climatology, Meterology and Oceanography* 3(1), 55–66. *https://doi.org/10.5194/ascmo-3-55-2017*
7. Calhoun, G. (2018) Block bootstrap consistency under weak assumptions. *Econometric Theory* 34(6), 1383–1406. *https://doi.org/10.1017/S0266466617000500*
8. Lahiri, S. N. (1999) Theoretical comparisons of block bootstrap methods. *The Annals of Statistics* 27(1), 386–404. *https://doi.org/10.1214/aos/1018031117*
9. Bühlmann, P., and Künsch, H. R. (1999) Block length selection in the bootstrap for time series. *Computational Statistics & Data Analysis* 31(3), 295–310. *https://doi.org/10.1016/S0167-9473(99)00014-6*
10. Bühlmann, P. (2002) Bootstraps for time series. *Statistical Science* 52–72. *https://doi.org/10.1214/ss/1023798998*

11. Hesterberg, T. C. (2015) What teachers should know about the bootstrap: Resampling in the undergraduate statistics curriculum. *The American Statistician* 69(4), 371–386. *https://doi.org/10.1080/00031305.2015.1089789*

12. Chernick, M. R., and Labudde, R. A. (2009) Revisiting qualms about bootstrap confidence intervals. *American Journal of Mathematical and Management Sciences* 29(3-4), 437–456. *https://doi.org/10.1080/01966324.2009.10737767*

13. Nevitt, J., and Hancock, G. R. (2001) Performance of bootstrapping approaches to model test statistics and parameter standard error estimation in structural equation modeling. *Structural Equation Modeling* 8(3), 353–377. *https://doi.org/10.1207/S15328007SEM0803_2*

14. Burch, B. D. (2012) Nonparametric bootstrap confidence intervals for variance components applied to interlaboratory comparisons. *Journal of Agricultural, Biological, and Environmental Statistics* 17(2), 228–245. *https://doi.org/10.1007/s13253-012-0087-9*

15. Puth, M.-T., Neuhäuser, M., and Ruxton, G. D. (2015) On the variety of methods for calculating confidence intervals by bootstrapping. *Journal of Animal Ecology* 84(4), 892–897. *https://doi.org/10.1111/1365-2656.12382*

16. Gonçalves, S., and White, H. (2005) Bootstrap standard error estimates for linear regression. *Journal of the American Statistical Association* 100(471), 970–979. *https://doi.org/10.1198/016214504000002087*

17. DiCiccio, T. J., and Efron, B. (1996) Bootstrap confidence intervals. *Statistical Science* 11(3), 189–228. *https://doi.org/10.1214/ss/1032280214*

18. Rice, J. A. (2006) *Mathematical Statistics and Data Analysis*. Cengage Learning, Boston

19. Efron, B., and Tibshirani, R. (1993) *An Introduction to the Bootstrap*. Chapman & Hall/CRC, Boca Raton. *https://doi.org/10.1201/9780429246593*

20. Carpenter, J., and Bithell, J. (2000) Bootstrap confidence intervals: When, which, what? A practical guide for medical statisticians. *Statistics in Medicine* 19(9), 1141–1164. *https://doi.org/10.1002/(SICI)1097-0258(20000515)19:9<1141::AID-SIM479>3.0.CO;2-F*

21. Efron, B. (1987) Better bootstrap confidence intervals. *Journal of the American Statistical Association* 82(397), 171–185. *https://doi.org/10.1080/01621459.1987.10478410*

22. R Core Team. (2022) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria

23. Canty, A. (2022) *boot: Bootstrap R (S-Plus) Functions*. R package version 1.3-28.1

24. Clopper, C. J., and Pearson, E. S. (1934) The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 26(4), 404–413. *https://doi.org/10.2307/2331986*

25. Scherer, R. (2018) *PropCIs: Various Confidence Interval Methods for Proportions*. R package version 0.3-0

26. Brown, L. D., Cai, T. T., and DasGupta, A. (2001) Interval estimation for a binomial proportion. *Statistical Science* 16(2), 101–133. *https://doi.org/10.1214/ss/1009213286*

27. Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York *https://doi.org/10.1007/978-0-387-98141-3*

28. Hofert, M., Kojadinovic, I., Mächler, M., and Yan, J. (2018) *Elements of Copula Modeling with R*. Springer, New York. *https://doi.org/10.1007/978-3-319-89635-9*

29. Geyer, C. J. (2011) Introduction to Markov chain Monte Carlo. In S. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng, editors, *Handbook of Markov chain Monte Carlo*, 3–48. CRC Press, Boca Raton *https://doi.org/10.1201/b10905*

30. Kreiss, J.-P. (1992) Bootstrap procedures for AR (∞)—processes. In *Bootstrapping and Related Techniques: Proceedings of an International Conference, Held in Trier, FRG, June 4–8, 1990*, 107–113. Springer, New York. *https://doi.org/10.1007/978-3-642-48850-4_14*

ABOUT STUDENT AUTHOR

Mathew Chandy is a senior majoring in both Statistics and Statistical Data Science, and he plans to graduate in the Spring of 2024.

PRESS SUMMARY

This simulation study evaluates the sample size necessary to estimate the mean, standard deviation, and lag-1 autocorrelation of a stationary time series using different block bootstrap confidence interval types. The results showed that percentile-based confidence intervals for the lag-1 autocorrelation may suffer from coverage deterioration as sample size is increased, motivating the authors to propose a new recentered percentile confidence interval which does not deteriorate in performance for greater sample sizes. The results also suggest that when using Student's t bootstrap confidence intervals, a sample size of at least 1600 may be sufficient to estimate the mean, whereas a sample size larger than 3200 may be necessary to estimate the standard deviation. The results additionally indicate that estimation of the lag-1 autocorrelation - using Student's t bootstrap confidence intervals - demands a sample size of at least 100 when the marginal distribution is standard normal and a sample size of at least 1600 when the marginal distribution is unit exponential.

Faculty Opinions of AI Tools: Text Generators and Machine Translators

Mahlet Yitages & Akie Kasai

Bass School of Arts, Humanities and Technology, The University of Texas at Dallas, Richardson, TX

https://doi.org/10.33697/ajur.2024.102

Students: malu.yitages@utdallas.edu, akie.kasai@utdallas.edu
Mentor: carie.king@utdallas.edu

ABSTRACT
Artificial Intelligence (AI) tools recently became a prominent concern in higher education classrooms. Many teachers have implemented the technology into their assignments, while others are strictly against this technology's use for assignments. Either way, students have found ways to use it in their academic careers. Though research on the power of AI in the workplace exists, research is lacking in its appropriate use in higher education. Universities need to define AI's role on campus and establish guidelines on how these tools may or may not be used and how faculty can recognize misuse, specifically related to academic integrity. This study aimed to determine how faculty view AI as a part of undergraduate literature, language, and linguistics programs. From the interview study, common themes emerged, including implementation, academic integrity, the human aspect of linguistics, and the future of AI writing tools. The faculty interviewed also stated that those in higher education must tread carefully through this strong intersection between technology and the arts to use AI responsibly, strategically, and ethically.

KEYWORDS: Artificial Intelligence (AI); Artificial General Intelligence (AGI); Linguistics; Higher Education; ChatGPT; Machine Translation; Academic Integrity; Ethics

INTRODUCTION

Artificial Intelligence (AI) writers and tools like ChatGPT (Generative Pretrained Transformer) have recently become part of daily discussions, as they appear extensively in news and televised entertainment.[1] For example, ChatGPT, created by OpenAI,[2] was integrated into South Park, a satirical adult-focused animation, when a portion of season 26 episode four was written using ChatGPT and several characters utilized it in the classroom to navigate various plot points.[3] In the episode, students used ChatGPT to write and submit essays without reviewing what ChatGPT said. In addition, the teacher graded the students' papers using ChatGPT because he was "burnt out" and he needed to work hard to spot the AI-generated text in students' essays.[3] By removing the human element in writing, this episode of South Park highlighted the problems of AI writing with a nonsensical and comedic conclusion for the episode. The characters resolved the problems caused by this AI technology by using ChatGPT to write a happy yet ironic ending that solved the issues and satisfied everyone, even if the ending made little sense plot-wise.

Realistically, the problem of AI tools in the classroom cannot be solved with a comedic *deus ex machina* plot device, so how can students and teachers deal with AI tools in the classroom?

Per the history of and literature on AI tools, the technology's rapid development has sparked a multitude of conversations. OpenAI introduced ChatGPT 3.5 to the world in December 2022, but the general public already knew of other AI technologies that OpenAI had released: e.g., earlier iterations of ChatGPT (GPT-2 in 2019 and GPT-3 in 2020) and the AI art generator called Dall-E in 2021. These AI tools were not necessarily new to the world, but GPT-3.5 was the first technology available for public use.[4] In March 2023, OpenAI released GPT-4, the latest version of this chatbot, which claims to process up to "25,000 words, about eight times as many as ChatGPT" and adds a new ability to respond to images that previous iterations of ChatGPT did not accomplish.[5] While conducting research to develop GPT-4, OpenAI made the model take tests like the "Uniform Bar Exam, the Law School Admission Test (LSAT), the Graduate Record Examination (GRE) Quantitative, and various Advanced Placement (AP) subject tests" and "scored at a human-level… if GPT-4 were a person being judged solely on test-taking ability, it could get

into law school—and likely many universities as well."[6] A close look at the research shows that GPT-4 tested poorly on the AP English Language and Composition and AP English Literature and Composition examinations, resulting in a score of 2, with 1 being the lowest score on the scale and 5 being the highest for AP examinations.[6] These accomplishments create concerns for students who need to take these tests—like the LSAT and GRE—that shape their career paths. Their scores can determine which university a graduate student attends, and AP subject tests can grant students potential college credit.[7] Knowing that AI tools are not experts at free-response examinations like the AP English exams shows that this model is not ideal for English essay writing. Considering these tests, faculty and testing institutions like College Board may need to consider how to ChatGPT-proof their examinations and prevent students from simply answering the questions with ChatGPT.

Undergraduate students and university faculty already use standard AI tools in increasingly technology-driven classrooms. Regarding AI tools, researchers propose solutions by considering student-, faculty-, and system-facing AI applications and discuss the implications for students, faculty, and entire institutions that ChatGPT and other tools can support.[8] Other researchers in the AI in Education (AIEd) field advise focusing on learning tools that utilize intelligent tutoring systems to personalize education to individual students and adapting to each student's learning style and using teacher tools that reduce workloads by automating certain procedures with automatic essay scoring systems and plagiarism detection systems (2023). Writing assistants like Grammarly[9] and Wordtune[10] claim to support both students and teachers to automate tedious tasks like checking grammar and correcting punctuation. However, these writing assistants cannot accomplish everything that a human proofreader can, especially related to context and commonly misspelled words. The question then is how to introduce these tools so students do not rely on these writing assistants, especially ChatGPT, for generating text and instead show students how to use these AI writers to improve their writing.

The credibility of ChatGPT and other AI tools may be clouded due to the companies behind the tools. *60 Minutes* interviewed technology specialists, including Brad Smith, President of Microsoft.[11] Smith mentioned the fears and hopes he had related to new AI technology in terms of helping people with Microsoft's implementation of AI in the Bing search engine.[11] However, the partnership between Microsoft and OpenAI aims to further profit companies and Microsoft is a stakeholder in AI tools like ChatGPT; their profit-driven goal is for both companies to "independently commercialize the resulting advanced AI technologies," so Smith highlights his aim of helping people and "creating advanced AI that benefits everyone" while simultaneously intending to profit off of this technology.[12] OpenAI is identified as a "capped-profit company" that combines ideas of both nonprofit and for-profit companies to "raise the capital we need to fulfill our mission without sacrificing our core beliefs," such as safety and broadly sharing the benefits with as many people as possible. Therefore, the partnership between Microsoft and OpenAI also aligns with OpenAI's mission: to create "artificial general intelligence (AGI) [that] benefits all of humanity, primarily by attempting to build safe AGI and share the benefits with the world," and they are aware that this multi-billion dollar investment from Microsoft allows OpenAI to continue researching and developing these AI tools for the sake of their mission.[13] Knowing who creates these tools and who funds these projects should encourage users to be wary of the intent behind these tools.

These AI tools are good at doing repetitive tasks over and over, but these tools can harm students in the learning process. Specifically addressing ChatGPT, *The Dallas Morning News* sponsored a panel of university professors and machine-learning experts who shared their knowledge in a panel discussion, "ChatGPT: Fact vs. Fiction."[14] In an interview, Assistant Professor of Computer Science Dr. Xinya Du stated, "[ChatGPT] can't be trusted with crunching numbers.... We can't really rely on the AI model to handle these calculations for important analysis" because AI models can regurgitate information that they already know, but when ChatGPT is "…asked 'what's 2 plus 2,' ChatGPT is only able to answer '4' because of previous answers on the internet."[15] AI tools cannot create new information; they recognize already established patterns and repeat words, much like a parrot can repeat a phrase.[15] It is unwise to rely on AI tools for every task; therefore, if students and faculty can access these tools, faculty need to teach students how, when, and where to use them.

AI tools have existed for decades, getting easier for non-technical users to access.[16] Specifically, AIEd has been researched since the 1970s,[8] with recent developments including Google Translate and Grammarly, both text-analysis tools that are a significant part of everyday use; now, integrating AIEd technology into the classroom is easier than expected.[17]

The potential of AI in classrooms is exciting, with the main benefit coming from personalized learning tools powered by AI for each student, automated repetitive tasks like grading and feedback, and instant feedback; however, those affiliated with arts and humanities fear the convenience of AI for other uses.[18] Whereas utilizing AI technology as assistive technology can help students with learning disabilities, "such students [depending] on technology like ChatGPT… encourages them to effectively outsource solutions rather than think for themselves or develop other ways of coping with their disabilities."[19] AI tools like ChatGPT are able to generate ideas and words; however, when prompted to build better arguments, "the different arguments provided by ChatGPT are overlapping and superficial... [and] serve better as idea prompts rather than ready-to-use content in writing."[20] Although AI tools can be implemented into the writing process and can generate "idea prompts rather than ready-to-use content in writing," as chatbots are designed to be an "intelligent conversational system" that students can engage, "students can neither converse with them freely on problems encountered during writing nor seek detailed feedback" like faculty feedback.[20] The limitations of plagiarism and "[generating] fake or misleading responses" cause experts to advise that ChatGPT should be "integrated with caution as... writing assistance in classrooms."[20] When students submit their work, "AI detectors like Turnitin and GPTZero suffer from false positives" and can wrongfully create suspicion of a student using AI tools, which could have "potentially devastating long-term effects."[21]

This study was created to analyze what STEM and humanities faculty say about AI tools in the classroom and if faculty should implement these AI tools, specifically writing and translation tools, in the classroom. AI tools like ChatGPT and GPT 4 have spread quickly among students, who need to be taught to use these tools to benefit them but without academic integrity issues.[22] Students who misuse these tools miss learning opportunities and submit subpar work that does not demonstrate the student's proficiency in a topic.[23] Without proper regulation and implementation, these methods will slip past professors and cause students to believe the work can be done for them.[22] Regarding AI tools that create written work, one discussion includes copyright within creative works.[24] Considering AI tools that translate one language to another, AI tools cannot translate abstract concepts, especially colloquial phrases and idioms that do not translate literally.[25] With the rise of AI tools in several career fields, the researchers decided to collaborate and merge the research projects together to further explore AI's potential use in the classroom.

METHODS AND PROCEDURES

With institutional review board (IRB) approval, we recruited and interviewed professors at a Carnegie tier-one research-oriented public university in the Southwestern US to obtain a unique perspective from professors at a STEM-focused school with a strong liberal arts program. The interviewed professors teach computer programming, literature, writing, language, translation, and communication. We utilized convenience sampling; the chosen professors were in proximity to one of the two researchers: that is, "the ones [professors] that are easier for the researchers to access."[26] Then, both researchers analyzed the interview narratives until we reached saturation, or "obtaining a comprehensive understanding by continuing to sample until no new substantive information is acquired"[26]. By using convenience sampling, the researchers obtained interviews with professors who teach at the same university, which may have created bias (as mentioned later in the study limitations section).

Through semi-structured interviews, we gathered demographic information and then asked the professors about AI writers and translators:

- Do you allow students to use technology in your classroom and to complete their work?
- What do you think of students using generative AI tools in the classroom?
- Have you used these generative AI tools personally?
- Have you implemented generative AI tools in the classroom/your curriculum?
- What is the ethical dilemma of AI implementation in education?
- What are your thoughts on AI tools in the classroom that we have not covered in other questions? Do you have anything on your mind that we have not covered yet?

We also asked questions based on the subject that the interviewee taught: computer science, literature and writing, and language and translation. Computer science professors were asked,

- How do you see AI developing in the future?
- What is the difference in the use of AI tools in STEM and humanities?
- Do you see AI progression as exciting or threatening?

- Do you believe there is potential for mass-produced fine-tuned models of AI assistants?

Linguistics, literature, and writing professors were asked,

- Do you see AI tools as helpful or harmful to students' learning experiences?
- How would you implement AI tools in the classroom?
- If we get to a point where we cannot determine if the text was written by a human or AI, does it matter who the author is if it is an engaging piece of literature?
- What is the benefit of using generative AI for writers? For faculty? For students?
- Have you come across AI writers in written assignments in your classroom?

Language and translation professors were asked,

- Would you implement machine translators in the classroom?
- What is the benefit of using generative AI for translators? For language learners?
- Do you think it is possible for human translators to work with machine translators?
- Should current translators and language learners rely on these new AI tools?
- Should there be a specific model for each language or have one universal translator? We also asked follow-up questions to expand the narrative within their responses.

We interviewed seven professors (see Table 1): three in literature, two in language and translation, and two in computer science. The sample included three female and four male professors. The professors differed in their ranks and roles; one was tenure track, three were tenured, and three were teaching intensive (non-tenured) professors.

Identifier	Gender	Job Title
Computer Science Professor 1	Male	Professor of Computer Science (tenured)
Computer Science Professor 2	Female	Assistant Professor of Computer Science (tenure track)
Language Professor 1	Female	Professor of Instruction and Director of Languages (teaching intensive)
Literature Professor 1	Male	Associate Professor of Instruction (teaching intensive)
Literature Professor 2	Female	Assistant Professor of Instruction (teaching intensive)
Rhetoric Professor 1	Male	Associate Professor of Rhetoric and Communication Studies (tenured)
Translation Professor 1	Male	Professor of Literature and Translation Studies (tenured)

Table 1. Demographics of professors who agreed to participate and were interviewed.

We conducted virtual interviews via Microsoft Teams from April 10 to April 28, 2023. One of the interviewers met with each professor to conduct one-on-one interviews and to reduce miscommunication and technical difficulties. After each interview, the interviewer saved the recording and transcription into OneDrive to allow both researchers access.

To analyze the interviewees' responses, we first noted common themes that became the center of our discussion. In addition to considering individual quotes from the interviews, we arranged the seven automatically generated speech-to-text Microsoft Teams

transcripts into a single document and removed the names and timestamps using regex expressions to automate the task. With that document, we analyzed the common terms mentioned in the interviews and removed filler words such as "like" or "um." Then, we experimented with a text analysis tool, Voyant Tools,[27] to analyze the interview transcripts and search for additional reoccurring keywords.

RESULTS AND DISCUSSION

The seven professors produced varied results throughout the study, but the three common themes were

(1) approaches to implementation,
(2) academic integrity and how it feeds into the originality of content, and
(3) the future of AI in the university classroom.

We address each of these themes separately.

Implementation

The professors interviewed were concerned about how AI tools could be effectively introduced into the classroom but the professors did not agree on implementation methods. What they did agree on was the ethical use of AI assistants. Literature Professor 1 referred to the tools as "pedagogically irresponsible." Rhetoric Professor 1 acknowledged, "I'm on a learning curve with this. I'm learning about it and hearing more and reading more... and my impression is that it can be both beneficial… and also detrimental."

Five of the professors hesitated to comment, acknowledging the added research that must be available before full integration. Faculty must be one step ahead of students in familiarizing themselves with these tools so they can be used ethically. Research shows that the rise in students using AI tools without authorization has caused scholars to be hesitant to include these tools in their teachings.[28] The lack of consideration for how students and faculty have used and will use technology in the classroom[8] has led to fearful responses, especially from those in humanities. Some professors suggest more oral examinations or in-class assignments to prevent plagiarism through AI writers.[29] By preventing students from using online tools, their work can be assessed without worry. Unfortunately, these methods are time-sensitive and create a greater sense of pressure—for students *and* faculty. Students struggle with these forms of assignments but students may have accommodations that prevent them from conducting these tasks under a specified time limit.

Literature Professor 1 sees avoiding these tools as detrimental to students' success. "It's not preparing students for their careers," Literature Professor 1 claims. "So… when you go out to work, [employers] don't say, 'Okay, sit down and write something in an hour." The workforce does not operate like a classroom, so students should be using tools that are currently used outside of the classroom.

The professors found potential benefits to adding these tools into their curricula. Literature Professor 1 idealized an interesting way to use machine learning models like Voyant in his Close Reading course(s) to narrow down terms in a text or use ChatGPT to "[find] what the machine looked for in the text." The misuse of ChatGPT has incited fear in academic administrators, but when used with caution, this tool and other AI tools can aid educators.[30] This is an optimistic viewpoint for AI in the humanities space regarding the analysis of long pieces of text, a time-consuming task, and can be one way to implement AI tools in the classroom.

Professors who are teaching language and translation are cautious in encouraging the use of machine translation, especially in classrooms. Language Professor 1 prohibits the use of AI tools like Google Translate in her Spanish classes and explains that, for students who want to improve their language skills, "using online translators to do your job… [students] won't be learning." Although she acknowledges that students want to earn high grades, she encourages students to make mistakes in the classroom. Students "need to know that making errors or mistakes is a good thing, because [students] can learn from them... If [students] use an online translator, they don't know what they are doing wrong." College students' fear of low grades can be explained because low GPAs can cause them to lose certain academic scholarships and other activities. This fear may entice students to take shortcuts to maintain high grades, possibly jeopardizing their academic integrity and their learning.

Language Professor 1 is a native Spanish speaker and usually can catch when students use tools like Google Translate to do their work. She noted that when she tested how these tools work, she noted that online translators could translate English to Spanish well, but for translating something she wrote in Spanish to English, it was "not that good," and she noted that if students were to write text in English, a language that they are fluent in, they don't have the same level of fluency in the target language they are translating into so it is "not the grammar they know…" and they "are not supposed to know those tenses or structures or how to express that" in the target language. When it comes to online translators, "English to... Spanish is pretty good, it's not perfect. But for things like subjunctive... or the difference between perfect and imperfect for the past tenses... [is] difficult for students [and even machine translators] to understand," meaning that the level of fluency in the target and source languages is not equivalent in the classroom. In addition, translating certain colloquial expressions can prove to be difficult because "[online translators] will make a literal translation." The translation professor even mentions that, when it comes to ChatGPT and AI tools that draw English text from the internet, it will work better in English than in other languages.

> [Approximately] 80% of the Internet is in English.... These are tools that are gonna [sic] work better on English texts than they are for texts coming from our language. Like what I work with, which is Romanian so these tools will be less impressive.... It's not that [ChatGPT] won't respond to [translation prompts,] it will respond to translation prompts, but they [AI tools] work better with text that has already been translated. (Translation Professor 1)

Translation Professor 1 noted that it could be possible to "accelerate the process of translation because people have already agreed on what the right translation for this text is," like in French, which uses long closing phrases on letters that are simply translated to 'yours sincerely' in English. When it comes to translating phatic expressions, or "phrase[s] that primarily serves to establish or maintain social relationships,"[31] it is possible to translate the same thing in countless different ways, depending on the context. Translating the same phrase over and over in the same way may be an efficient way to translate, especially for simply getting the meaning across to someone; however, a human translator can make the creative decision to translate a certain phrase differently based on context and speaker and other reasons.

Concerning implementation, students must learn how to navigate technology in this rapidly evolving world, especially with the introduction of AI. In a panel on ChatGPT, Associate Dean of Research and Creative Technologies Dale MacDonald said, "It is becoming clear that it's important that students use it. ... Our students are going to have to have this literacy."[14] With students already becoming familiar with AI writers, professors must do the same to create boundaries with such tools. Implementation within higher education courses seems clear, but faculty and student use of these tools must be determined. The task of implementing AI tools in linguistics is a difficult feat. The concerns about academic integrity and originality of text have caused experts to consider the ethics behind AI assistants.[32] Most consumers agree that AI must be regulated, but the issue is who will be responsible and how these tools will be moderated.[32]

If considered with an open mind, these tools can improve the expectations placed on students. If these models can conduct "peer reviews" or edits before a student runs through a text, then the student's work can be elevated. However, with more tools comes higher expectations of students. Literature Professor 2 especially expressed concerns of potential added pressure to students who already cannot manage their academic workloads: "…my concern is now what are going to be the increased expectations on people's time and… output." Although these tools help if fully implemented and established as the standard, instructors must consider what more will be required of students. When questioned about how faculty should respond to a possible implementation, Dr. Du stated, "We can come up with questions that are more challenging."[14] There has been a natural comparison to the use of calculators:

> You know, calculators were invented. You might give homework where you actually have to do the calculations, but I guess today nobody gives those kinds of assignments…. And so now the game changes in some sense that now the nature of homework has to change. (Computer Science Professor 1)

Students are taught when and how to use calculators, which mathematical problems can be solved with calculators, and which equations are elementary enough to use a calculator. AI tools must be treated similarly; students must be taught *how* to use AI tools to ensure that they still learn to read and communicate effectively.

Computer Science Professor 2 went in depth about how these tools cannot be compared directly to calculators: "So it's not the task that we're asking you to do… It's the fact that you've gone through the process of doing the task. And that's the learning experience we're looking for." Without proper observation of the use of AI tools in students' work processes, they can lose the benefits of the learning experience. Allowing students to use machine learning tools, which help with editing, can speed up the writing process by cutting hours of reviewing (e.g., tools like Grammarly or spell check/autocorrect [Translation Professor 1]). Students must learn what syntactical edits to look for before using these tools. For AI assistants to be useful, the student must be smarter than the computer.

Academic Integrity

The professors who we interviewed sometimes disagreed on the use of tools and whether implementing AI tools in the classroom would benefit students in higher education. However, they shared the same primary concern: academic integrity. They see misuse of AI writing and translation tools as a threat to students' education, and several stated that they fear the implementation of AI use could deter students from working and learning to their full abilities.

Numerous tools in the past have caused professionals to assume their only use would be for cheating. The example of calculators used by Computer Science Professor 1 helps us understand that any tool produced can be used for good or bad, depending on its user. It has become the job of faculty in higher education to monitor how and when students are using these tools. Professors cannot control whether their students use AI assistants to cheat, but they can familiarize themselves with the pattern-based writing seen with these assistants. Machines still do not have their own voice; the brainpower that is needed to apply context to writing has not been achieved by machine[s] yet. "Everybody has a unique voice," Literature Professor 1 said. "Just like everyone has a unique fingerprint." The fear is that one day, machines will be able to think like humans do; though the technology is not there yet, it can surely come soon.

The blurred lines between citation guidelines within computer science and linguistics have confused what constitutes a breach of academic integrity. A "gray area" exists for computer science citations:

> To me… the main difference is that in computer science… you're either right or you're wrong…. Whereas I think the beauty of the humanities is if you have five people answer that question, you probably get five completely different answers. But they're all right… (Computer Science Professor 2).

AI writing assistants bridge the gap between computer science and linguistics. Because experts in computer science have familiarized themselves for years with this technology, this community experiences more amazement. In contrast, linguist faculty were alarmed by advancements in AI like ChatGPT. Computer Science Professor 2 explained that, although the tool can supply code found within websites like GitHub, it only provides a faster way for students to find the source, like a search engine. "It's a way to access information," she said. "Maybe a more convenient way than Google, and it feeds you the results in a conversational way." ChatGPT cannot be used in the same way for those studying linguistics; Rhetoric Professor 1 states that when it comes to using AI assistants, a "fine line" exists for what constitutes cheating. "It's like any other tool; it can be used for good purposes and for not-so-good purposes." Because these evolved tools serve similar goals as older, common tools, students and faculty should approach newer tools with the same attitude of acceptance paired with careful consideration.

The spike in student usage of ChatGPT in early 2023 brought a heavy weight on university faculty across the world.[33] Because of the rise in the popularity of AI assistants, the "concept of plagiarism" has been brought into question.[8] Though these tools were meant to be helpful, the extensive range of skills these tools have can hinder students' learning experience. "Unfortunately, [shortcuts] are going to encourage and maybe even incentivize…academic dishonesty and breach academic integrity." Rhetoric Professor 1 was most concerned with how machine learning technology could impact his advanced editing classes. Like most humanities faculty, he fears these tools may be abused for their power. A large concern of experts in higher education, especially when referring to ChatGPT, concerns students who fabricate citations.[34] With the tools previously mentioned, students can create sources that "support" their argument rather than investigating and referencing true sources.

In addition to the concern of unethical academic use, businesses are scrambling to organize copyright laws for AI-generated content. AI art and music concern artists, as many do not know who to credit for these works. "…There's incredible beauty in a

human being's ability to work on a craft and develop it and improve upon it," stated Literature Professor 2, who emphasized the human touch needed in creation. "I still feel like there's incredible beauty in a human being's ability to work on a craft and develop it and improve upon it." The issue is whether the authenticity of works will still be valued as these tools grow in popularity. She also presented the dilemma of withholding information on how a creative work was made. She said creators must be upfront about AI usage: "Then I think they'll be less [anxious] about it… It was made by a machine, but I know that it's made by a machine." Though the work of a machine cannot live up to the text humans can produce, knowing who or what should be credited is vital.

Future of AI

Although the five humanities professors were skeptical about AI's future in education, the two computer science professors voiced that these tools could improve students' education. AI will not be dismissed,[35] so university faculty are considering what changes universities must make. "People will absolutely be adapting them," Computer Science Professor 1 stated. "Fine-tuning them provides what is called in-context learning."

On the ChatGPT panel (previously introduced), Dale MacDonald reassured his audience that their jobs were not at risk. "AI is not going to take your job," MacDonald stated.[15] "A human that can use AI is going to take your job."[14] Computer Science Professor 2 reiterated this statement in her interview: "I think no matter what, there needs to be some human who's babysitting the AI system." At this time, we cannot say or do much about the future of technology, but we can try to get ahead by setting standards for what we have today. As Computer Science Professor 1 said, "Since [the] domain is narrow, [researchers] cannot reason outside of this domain." With the possibility of mass-produced fine-tuned AI assistants, researchers worry about data collection. This concern was explored further at the ChatGPT panel, when one professor advised, "It's always a concern that anything you put on the internet could be collected (Good).[14]

With what we have seen with AI, instructors have hopes and concerns for the future of these AI tools. There is talk of possible fine-tuned models made for specific skill sets. When OpenAI began developing ChatGPT, AI trainers were required to create these fine-tuned models.[2] Specialization is one possible positive outcome of machine-learning tools, but many people fear that AI will become "too powerful." The future of AI is vast but talk of AGI (Artificial General Intelligence) has grown.[36] Open AI's plan for AGI is to create "AI systems that are generally smarter than humans."[36] Scholars are skeptical of AI and even more concerned about stronger AI.

CONCLUSIONS: *STUDY LIMITATIONS, FUTURE RESEARCH, AND STUDY VALUE*

The main limitations of our study include the length of time for data collection, the number of professors interviewed, and a concern of bias due to the overflow of media coverage on AI at the time of our study.

We had only three weeks to conduct interviews, which also limited the number of professors we were able to interview. The researchers' and professors' schedules were full toward the end of the semester, and this study was conducted within a senior-level research and writing course with a deadline. The study was centered around discussions with university experts in computer science and in linguistics and writing. Universities that are focused heavily on engineering and science do not face an insurmountable connection between students who study humanities and those who study technology.[37] This study focused more on the liberal arts and literature, as both researchers were majoring in literature with a concentration in rhetoric and communication. However, both researchers briefly studied computer science and software engineering respectively, before changing majors, so both researchers had a basic knowledge of the technology field.

In addition, the researchers focused only on faculty in the humanities and computer science departments for their sample of convenience.[26] Because media focused on ChatGPT the professors were aware of the tool. This technology evolves quickly and news about this new technology surfaces regularly. It is impossible to encompass all the development that has happened in the last few months (since the research was completed).

In future iterations of this study, researchers should seek a balanced ratio of humanities and computer science professors, as well as scholars in business, political science, and other interdisciplinary studies. Every professor interviewed knew of ChatGPT, but

many had not realized how many of the tools they use are considered machine learning; we anticipate value in learning how faculty in a variety of fields view AI tools and what future they see with these tools. Furthermore, future research could consider a student perspective, especially once instructors are integrating or prohibiting AI into their classes and programs.

Professors have become more interested than fearful as they have learned more about machine learning and what these tools can do to improve education. Faculty understand that students cannot neglect their linguistic skills for the sake of convenience but also that these tools can help students create better texts. Rather than trying to keep technology and humanities separate, students should be encouraged to use machinery to help them create something unique. These tools should not be used as a starting point for written works but can be used to improve them.

It is important to note that AI is a developing technology and our study was conducted in April 2023, providing a brief snapshot of how university professors viewed these developing tools at the time. Certain automated tasks like spellcheck are already implemented in our everyday lives; however, the academy must view these tools with caution and avoid developing an overreliance on them because of the convenience and novelty of these tools. Even automated tasks like spellcheck need to be checked by a human.

ACKNOWLEDGEMENTS

The authors thank Dr. Carie S. Tucker King for her support and supervision throughout the experiment and writing process. She believed that this paper could be published in the first place and this study and publication would not have happened without her help. We also would like to thank the seven professors from the Bass School of Arts, Humanities, and Technology and the Erik Jonsson School of Engineering and Computer Science for their participation.

REFERENCES

1. Thorpe, H. (2023, Jan 26) Chat GPT is fun, but not an author, *Science*, 379 (6630), *https://www.science.org/doi/full/10.1126/science.adg7879* (accessed Sep 2023)
2. OpenAI. (2022) *Introducing ChatGPT*, *https://openai.com/blog/chatgpt (*accessed Sep 2023)
3. Parker, T. (Writer), ChatGPT (Writer), and Stone, M. (Director). (2023, Mar 8) "Deep Learning" (Season 23, Episode 4), *South Park,* Comedy Central.
4. Kay, G. (2023, Feb 1) The history of ChatGPT creator OpenAI, which Elon Musk helped found before parting ways and criticizing, *BI, https://www.businessinsider.com/history-of-openai-company-chatgpt-elon-musk-founded-202212* (accessed Sep 2023)
5. Derico, B., and Kleinman, Z. (2023, Mar 14) OpenAI announces ChatGPT successor GPT-4, *BBC, https://www.bbc.com/news/technology-64959346* (accessed Sep 2023)
6. OpenAI. (2023, Mar 15) *GPT-4 technical report*, *https://doi.org/10.48550/ARXIV.2303.08774* (accessed Sep 2023)
7. Edwards, B. (2023, Mar 14) OpenAI's GPT-4 exhibits "human-level performance" on professional benchmarks, *Ars Technica, https://arstechnica.com/information-technology/2023/03/openai-announces-gpt-4-its-next-generation-ailanguage-model* (accessed Sep 2023)
8. Rudolph, J., Tan, S., and Tan, S. (2023) ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?, *J Appl Learn Teach*, *6*(1), *https://doi.org/10.37074/jalt.2023.6.1.9* (accessed Sep 2023)
9. Chowdhury, R. (2023) The future of AI lies in augmented intelligence, *Grammarly*, *https://www.grammarly.com/blog/augmented-intelligence* (accessed Sep 2023)
10. Goshen, O. (2023) Generative AI will assist writers, not replace them in 2023, *Wordtune*, *https://www.wordtune.com/blog/future-ai-writing* (accessed Sep 2023)
11. Stahl, L. (2023, Mar 6) 60 Minutes—ChatGPT: Artificial Intelligence, chatbots and a world of unknowns, *YouTube, https://youtu.be/1wzPr4cUoMQ* (accessed Sep 2023)
12. Microsoft Corporate Blogs. (2023, Jan 23) Microsoft and OpenAI extend partnership, *Official Microsoft Blog*, *https://blogs.microsoft.com/blog/2023/01/23/microsoftandopenaiextendpartnership* (accessed Sep 2023)
13. OpenAI Blog. (2023, Jan 23) *OpenAI and Microsoft extend partnership*, *https://openai.com/blog/openai-andmicrosoft-extend-partnership* (accessed Sep 2023)
14. Good, J. (2023, Mar 29) Technology experts separate ChatGPT fact from fiction at forum, *UTD News Center, https://news.utdallas.edu/campus-community/chatgpt-forum-2023* (accessed Sep 2023)

15. Castillo, A. (2023, Mar 23) UTD scientists clear misconceptions, raise concerns about viral AI chatbot, *The Dallas Morning News*, *https://www.dallasnews.com/news/2023/03/23/utd-scientists-clear-misconceptions-about-viral-aichatbot-chatgpt-but-raise-concerns* (accessed Sep 2023)
16. Alam, A. (2021) Possibilities and apprehensions in the landscape of artificial intelligence in education, *2021 Intl Conf Comput Intelligence Comput Appl*, 1–8.
17. Alam, A. (2022) Employing adaptive learning and intelligent tutoring robots for virtual classrooms and smart campuses: Reforming education in the age of artificial intelligence, *Adv Comp Intelligent Technol*, 395–406.
18. Shonubi, O. (2023, Feb 21) All in the classroom: Pros, cons, and the role of EdTech companies, *Forbes*, *https://www.forbes.com/sites/theyec/2023/02/21/ai-in-the-classroom-pros-cons-and-the-role-of-edtechcompanies/?sh=17edf5ccfeb4* (accessed Sep 2023)
19. Smith, G. (2023) Chatgpt in the classroom: Friend or foe?, *ITNOW*, *65*(2), 46–47,
20. Su, Y., Lin, Y., and Lai, C. (2023) Collaborating with ChatGPT in argumentative writing classrooms, *Assessing Writing*, *57*, 100752, *https://doi.org/10.1016/j.asw.2023.100752 (*accessed Sep 2023)
21. Fowler, G. (2023, Aug 14) What to do when you're accused of AI cheating, *TWP, https://www.washingtonpost.com/technology/2023/08/14/prove-false-positive-ai-detection-turnitin-gptzero* (accessed Sep 2023)
22. U. S. Department of Education. (2023), Artificial Intelligence and Future of Teaching and Learning: Insights and Recommendations, Washington, DC: Office of Educational Technology.
23. Ahmad, S., Han, H., Alam, M., Rehmat, M., Irshad, M., Arrańo-Muńoz, M., and Ariza-Montes, A. (2023) Impact of artificial intelligence on human loss in decision making, laziness and safety in education, *Huma Soc Sci Comm*, 1–14.
24. Guadamuz, A. (2017, May) Artificial Intelligence and copyright, *WIPO Magazine*, *https://www.wipo.int/wipo_magazine/en/2017/05/article_0003.html* (accessed Sep 2023)
25. Ordorica, S. (2023, Jun 5) Comparing and contrasting AI and human translation, *Forbes*, *https://www.forbes.com/sites/forbesbusinesscouncil/2023/06/05/comparing-and-contrasting-ai-and-humantranslation/?sh=759790c932fd* (accessed Sep 2023)
26. Etikan, I., Abubakar Musa, S., and Sunusi Alkassim, R. (2015) Comparison of convenience sample and purposive sampling, *Am J Theoret Appl Stat*, 5(1), 1–4, *https://doi.org/10.11648/j.ajtas.20160501.11* (accessed Sep 2023)
27. Voyant. (2003) *Voyant Tools*, *https://voyant-tools.org* (accessed Sep 2023)
28. Jimenez, K. (2023) 'This shouldn't be a surprise' The education community shares mixed reactions to ChatGPT, *USA Today*, *https://www.usatoday.com/story/news/education/2023/01/30/chatgpt-going-banned-teachers-sound-alarmnew-ai-tech/11069593002* (accessed Sep 2023)
29. Huang, Kalley. (2023, Jan 16) Alarmed by A.I. chatbots, universities start revamping how they teach, *TNYT*, *https://www.nytimes.com/2023/01/16/technology/chatgpt-artificial-intelligence-universities.html* (accessed Sep 2023)
30. Roose, K. (2023, Jan 12) Don't ban ChatGPT in schools. Teach with it, *TNYT*, *https://www.nytimes.com/2023/01/12/technology/chatgpt-schools-teachers.html* (accessed Sep 2023)
31. Leonoudakis, K. (2023, Apr 18) Translating phatic expressions, *Katrina Leonoudakis Blog, https://katrina-ltranslation.online/2023/04/18/translating-phatic-expressions* (accessed Sep 2023)
32. Pazzanese, C. (2020) Ethical concerns mount at AI takes bigger decision-making role in more industries, *The Harvard Gazette*, *https://news.harvard.edu/gazette/story/2020/10/ethical-concerns-mount-as-ai-takes-bigger-decisionmaking-role* (accessed Sep 2023)
33. U. C. Berkeley. (2023) Understanding AI Writing Tools and their uses for teaching and learning at U.C. Berkley, *Berkeley Center for Teaching and Learning*, *https://teaching.berkeley.edu/understanding-ai-writing-tools-and-theiruses-teaching-and-learning-uc-berkeley* (accessed Sep 2023)
34. Welborn, A. (2023, Mar 9) ChatGPT and fake citations, *Duke University Library*, *https://blogs.library.duke.edu/blog/2023/03/09/chatgpt-and-fake-citations* (accessed Sep 2023)
35. Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel T., Stadler, M., and Kasneci, G. (2023) ChatGPT for good? On opportunities and challenges for large language models for education, *Learning and Individual Differences*, 1–9.
36. Knight, W. (2023) Some glimpse AGI in ChatGPT. Others call it a mirage, *Wired*, *https://www.wired.com/story/chatgpt-agi-intelligence* (accessed Sep 2023)

37. Strauss, V. (2017, Oct 18) Why we still need to study the humanities in a STEM world, *TWP*, *https://www.washingtonpost.com/news/answer-sheet/wp/2017/10/18/why-we-still-need-to-study-the-humanitiesin-a-stem-world* (accessed Sep 2023)

ABOUT THE STUDENT AUTHORS

Akie Kasai and Mahlet Yitages are pursuing Bachelor of Arts degrees in Literature, with Concentrations in Communication and Rhetoric from The University of Texas at Dallas. Kasai will graduate in December 2023, and Yitages will graduate in May 2024. They began their studies in computer science and software engineering, respectively.

PRESS SUMMARY

In the higher education setting, artificial intelligence (AI) tools recently became a prominent concern for university faculty. With AI rapidly gaining popularity within higher education around the world, the topic of ethics in technology has also been brought to the forefront of research. In the university setting, faculty are becoming more aware of these AI tools that students have access to. There is a spectrum of responses regarding these tools, from some actively implementing AI tools in their curriculum to some prohibiting these tools at all in the classroom. Either way, students are learning in an environment where these tools are becoming more prevalent, so our study aimed to hear from humanities and STEM faculty to determine how to implement AI tools, if at all, in the classroom. Based on seven interviews conducted in April 2023 with faculty from the humanities department and the engineering & computer science department, concerned with the rise of AI tools in their classrooms, the paper concludes that if these AI tools are to be implemented in the classroom ethically, they should be used as helping tools like how a calculator is used in the classroom and not as a starting tool for AI tools to do the student's work for them.

The Predicted Structure of a Thermophilic Malate Synthase

*Shaelee Nielsen, Jantzen Orton, & Bruce R. Howard**

Department of Chemistry and Physics, Southern Utah University, Cedar City, UT

https://doi.org/10.33697/ajur.2024.103

Student: shaeleenielsen@gmail.com, jantzen.orton@gmail.com
*Mentor: howard@suu.edu**

ABSTRACT
This project aims to solve the structure of the crenarchaeal *Sulfolobus acidocaldarius* enzyme malate synthase. Other known malate synthase enzymes have been found to require a magnesium ion in the active site to carry out catalytic activities, but a study reported that *S. acidocaldarius* malate synthase does not require magnesium. This suggests a novel mechanism for this enzyme. Additionally, the mature *S. acidocaldarius* protein is approximately 100 residues larger than any other structurally characterized malate synthase. It has also been reported to form a dimer, while previously solved structures have only displayed monomeric, trimeric, and hexameric arrangements. We plan to determine the structure experimentally. However, major advances in the accuracy of protein structure prediction were made recently by AlphaFold, an artificial intelligence system developed by DeepMind, which has revolutionized the field and has largely solved the protein folding problem. A similar AI system, RoseTTAFold, developed by David Baker's lab at the University of Washington, has been made publicly available. Here, we report our analysis of the structure of this protein, predicted using both of these algorithms and of a predicted structural model for the dimeric form of the enzyme using ClusPro. Our results strongly support a conserved catalytic mechanism requiring magnesium, which is common with all previously solved malate synthase isoforms.

KEYWORDS
Glyoxylate Cycle; Malate synthase; Protein Prediction; Thermophile; *Sulfolobus acidocaldarius;* Magnesium; AlphaFold; RoseTTAFold

INTRODUCTION

The glyoxylate cycle is a metabolic pathway that relies on several citric acid cycle enzymes along with two additional enzymes, isocitrate lyase and malate synthase, to synthesize citric acid cycle intermediates from two-carbon units.[1] Evidence of this pathway was first discovered in Bacteria (*E. coli*),[2] but the cycle's enzymes have since been identified in Eukarya and Archaea, as well.[3-5] Operation of this cycle is essential for many microorganisms to grow on a media where acetate is the sole carbon source.[6] Furthermore, the glyoxylate cycle contributes to survival in the pathogenic organisms *Candida albicans* and *Mycobacterium tuberculosis*, once they are engulfed by a host macrophage.[7, 8]

In organisms employing the glyoxylate cycle, an isocitrate molecule is cleaved to form succinate and the two-carbon compound glyoxylate by isocitrate lyase. In step two, malate synthase catalyzes the Claisen condensation of glyoxylate and an incoming acetyl-CoA to form a malyl-CoA intermediate that is then hydrolyzed to malate and CoA. In this way, the two decarboxylation steps of the citric acid cycle are bypassed, which generates additional precursors for amino acid and carbohydrate biosynthesis.

Previous efforts have identified four isoforms of malate synthase but only three isoforms have been structurally characterized.[9] Originally, two of these were distinguished by whether growth was observed on acetate (MSA) or glycolate (MSG) in *E. Coli*, which contains both.[10] All solved structures in bacteria are one of these two isoforms: MSA or MSG. Although both isoforms, A and G, exist as monomers in currently determined structures, their size and amino acid sequence vary significantly. Members of isoform A

consist of ~530 amino acid residues, whereas those from isoform G consist of ~730 residues; the structurally conserved regions of the sequences retain an 18% identity.[11] MSA is predicted to form oligomers in some cases.[12, 13]
The two remaining isoforms have only been identified in Archaea at present. The first reported example of an archaeal malate synthase to be purified came from *Haloferax volcanii*,[5] a halophile that grows in the bottom sediment of the Dead Sea at high salt concentrations.[14] Interestingly, this protein does not closely resemble MSA or MSG sequences and is only 433 amino acid residues.[15] Due to these differences and the discovery of other halophilic malate synthases through genome sequencing,[16] this class of enzyme was denoted as MSH.[17] Data from gel-filtration chromatography and crystal structures revealed that isoform H exists in a trimer/hexamer equilibrium, and in comparison to MSA and MSG, exhibits a deletion at the N-terminus and a truncated C-terminal domain. Despite these variations, the geometry of the active site and catalytic mechanism were conserved.[9]

The central protein fold that contains the active site in the aforementioned isoforms is a $(\beta/\alpha)_8$ (TIM) barrel. A crystal structure of the glyoxylate complex in *E. coli* MSG revealed that Asp 631 donated from a C-terminal domain, and Arg 338 within the TIM barrel are both conserved to provide acid-base chemistry; Asp 631 deprotonates the methyl group of acetyl-CoA, which subsequently forms an enolate that is stabilized by the positive charge carried by Arg 338. Two other critical residues, Glu 427 and Asp 455, coordinate an essential Mg^{2+} ion into place, thereby polarizing the glyoxylate substrate for nucleophilic attack. The positive charge carried by Mg^{2+} (cooperating with Arg 338) also helps stabilize the oxyanion that is formed from the attack of the enolate intermediate.[18] **(Figure 1)**

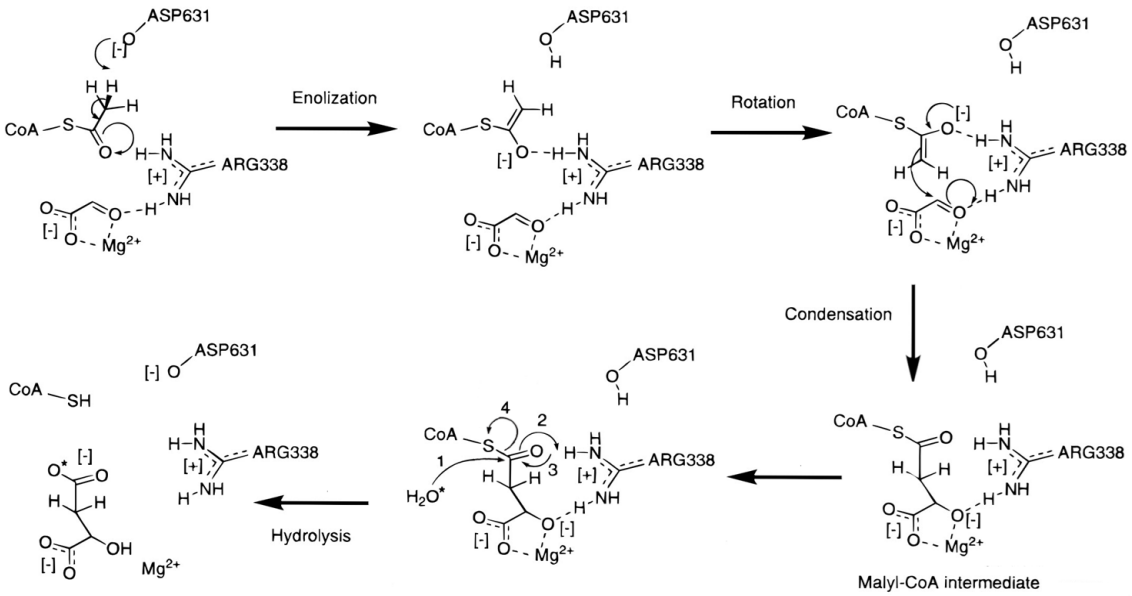

Figure 1. A currently proposed enzymatic mechanism for malate synthase with *E. coli* MSG numbering.[18]

The fourth isoform of malate synthase, found in crenarchaeal species like *Sulfolobus acidocaldarius*, has yet to be structurally solved. This organism is a thermoacidophile that grows optimally at 75-80 °C and a pH of 2-3, typically found in acid thermal soils or acid hot springs.[19] Genome sequencing revealed that the malate synthase protein is 824 residues in length,[20] about 100 residues longer than MSG. In contrast to the other structures, it was reported to function without Mg^{2+} and to form a functional homodimer.[21] The significant difference in cofactor dependence suggests that the enzyme may employ a novel catalytic mechanism.

In order to investigate this unique type of malate synthase, its catalytic requirements, and its evolutionary relationship compared to previously characterized isoforms, we have undertaken the present study. Simultaneously, our study led us to explore the recent advancements in protein structure prediction through the application of artificial intelligence technology. AlphaFold, developed by DeepMind, has revolutionized structural biology by demonstrating accurate predictions of protein structures comparable in quality to experimentally-determined structures.[22] RoseTTAFold, is a similar system that was developed by David Baker's lab at the University of Washington.[23] Both systems were released in 2021. Furthermore, the ClusPro web server uses computational methods

to sample billions of potential protein-protein conformations.[24-27] These results are further refined through energy minimization calculations that produce the most likely models for a given interaction. These systems were implemented into the study design in order to explore their potential application to this enzyme.

METHODS AND PROCEDURES

Analysis of the Predicted Structures

The protein sequence was retrieved from the UniProt database (Identifier # Q4J6V5) and submitted to the RoseTTAFold public server found at https://robetta.bakerlab.org/ for structural predictions. The structure output by the server was downloaded as a PDB file and was viewed using PyMOL.[31] Analysis of the structure allowed us to identify a TIM barrel that appeared homologous to other barrels that contain the active site in previously determined structures of malate synthase.

In order to conduct a more detailed comparison, representative structures for each one of the known malate synthase isoforms were downloaded from the Protein Data Bank (PDB):[32] MSA from *Bacillus anthracis* (3CUX), MSG from *Escherichia coli* (1D8C), and MSH from *Haloferax volcanii* (3OYZ). The following four conserved catalytic residues were identified inside each TIM barrel: an arginine to act as a bidentate cation which bridges the two substrates during the course of the reaction, an aspartate that acts as the catalytic base, and another aspartate cooperating with a glutamate residue to chelate the required cofactor. **(Figure 1)**

To compare the spatial positioning of the TIM barrel surrounding these key residues, protein alignment and overlays were performed. The MSG isoform from *Escherichia coli* (1D8C) served as the stationary molecule to which all overlays were made. We isolated the 3D coordinates of the C-alphas from each of the four key catalytic residues in each isoform and created PDB files containing solely those coordinates.

Using the PyMOL alignment tool, each of these sets of coordinates were aligned onto the fixed coordinates for these four C-alpha atoms of 1D8C. The LSQ algorithm in the COOT (Crystallographic Object-Oriented Toolkit) software was used to overlay each protein onto its corresponding four-coordinate aligned file.[28,29] These overlaid structures were saved and viewed simultaneously in PyMOL, allowing for direct structural comparison of the active sites.

For each isoform and the RoseTTAFold model, the distances were measured between the C-alphas of these residues as (A) aspartate base to arginine, (B) Mg^{2+}-chelating aspartate to arginine, (C) Mg^{2+}-chelating aspartate to glutamate, and (D) Mg^{2+}-chelating aspartate to the aspartate base **(Figure 3, Table 1)**.

The AlphaFold predicted structure was retrieved through the UniProt database (Identifier # A0A0U3GUU1). As with the other structures, the coordinates of the four catalytic residue C-alphas in the AlphaFold structure were aligned on those in 1D8C and the respective distances were measured. Additionally, overlays of the AlphaFold and RoseTTAFold structural models were compared and analyzed using PyMOL.

Prediction of a Dimeric Form

The ClusPro protein-protein docking server at https://cluspro.bu.edu/ was accessed using the "Dimer Classification" feature to predict potential dimeric forms of the RoseTTAFold model.[30] The server output 26 potential structures according to the "balanced" ranking option. PyMOL was used to view and draw conclusions from these dimeric models.

RESULTS AND DISCUSSION

Conserved Mechanism

The overlay of MSG, MSA, MSH, and the RoseTTAFold model indicates that the $(\beta/\alpha)_8$ (TIM) barrel domain is conserved across the various isoforms **(Figure 2)**.

At an atomic level, the distances between catalytic residues are also conserved. The distances between the key residues in the active site are very similar across each representative isoform, including both of the predicted structures for *S. acidocaldarius* **(Figure 3,**

Table 1). This further suggests that the functions of the catalytic arginine, base, and both Mg^{2+}–coordinating residues are maintained in the fourth isoform of malate synthase from *S. acidocaldarius*. Hence, we propose the following functions in *S. acidocaldarius:* Arg-214 acts as a bidentate cation, Asp-600 acts as the catalytic base, and both Asp-326 and Glu-298 chelate the suggested cofactor.

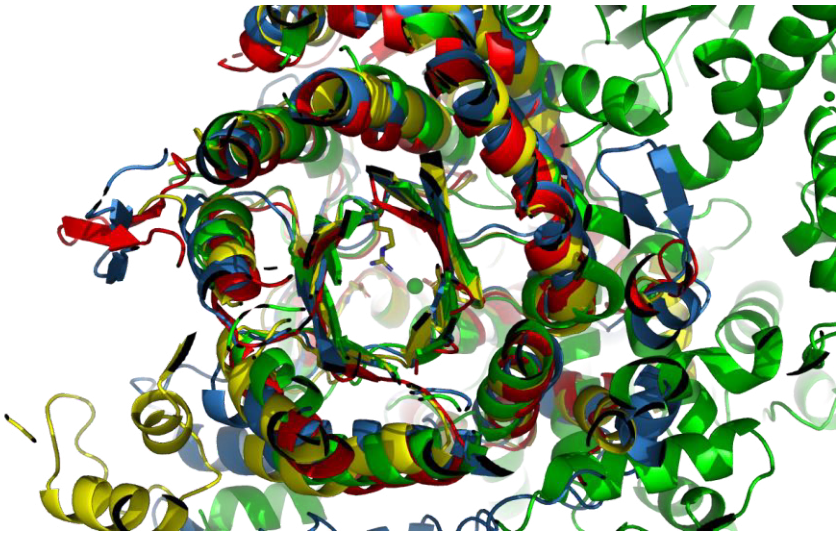

Figure 2. Overlay of MSA (red), MSH (green), and the RoseTTAFold model (blue) onto *E. coli* MSG (yellow).[31]

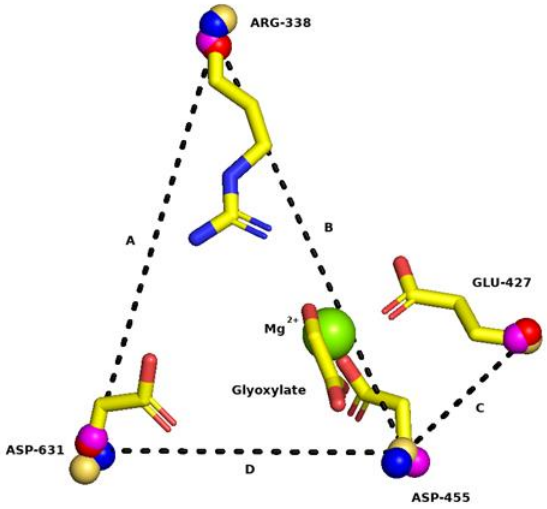

Figure 3. Relative position of Cα-atoms of key catalytic residues and relative distance measurements shown for the RoseTTAFold model (blue spheres). Also shown are the Cα-atoms for MSA (red spheres), MSH (yellow spheres), and the AlphaFold model (purple spheres) aligned onto those of MSG.[31]

Distance (Å)	MSG	MSA	MSH	RoseTTAFold	AlphaFold
A	13.9	13.4	15.1	14.3	13.3
B	14.7	14.5	15.1	15.2	15.0
C	5.5	5.8	5.6	5.8	5.4
D	10.1	10.1	10.2	9.6	10.4

Table 1. Relative distances between Cα-atoms of key catalytic residues shown in Angstroms (Å) for each isoform and the predicted model.[31]

The conserved $(\beta/\alpha)_8$ (TIM) barrel domain and the conserved positions of key catalytic residues within the barrel provide strong evidence for a conserved catalytic mechanism in the malate synthase of *S. acidocaldarius*. A proposed mechanism for malate synthase in previously solved structures of MSG, MSA, and MSH requires magnesium in the active site (**Figure 1**). The conserved TIM-barrel fold and positioning of these key catalytic residues include the magnesium-coordinating groups. This calls into question the report that this isoform did not require magnesium for activity.[21] It is important to remember that the structures used in the analysis are only predictions, so results cannot be confirmed until the structure of the protein is determined experimentally.

Predicted Dimer

The "Dimer Classification" feature from the ClusPro server using the balanced energy function produced the top 26 dimer predicted arrangements using the RoseTTAFold model.[30] Using PyMOL, all of these dimer predictions were viewed simultaneously by superimposing one subunit of each dimer **(Figure 4)**. This allowed us to compare the relative positions of the predicted interfaces. Each of the top 26 predicted dimer interfaces are found within one of 3 main locations on the RoseTTAFold model. The green-yellow-tan cluster contains 21 predictions, the blue cluster contains 3 predictions, and the red cluster contains 2 predictions.

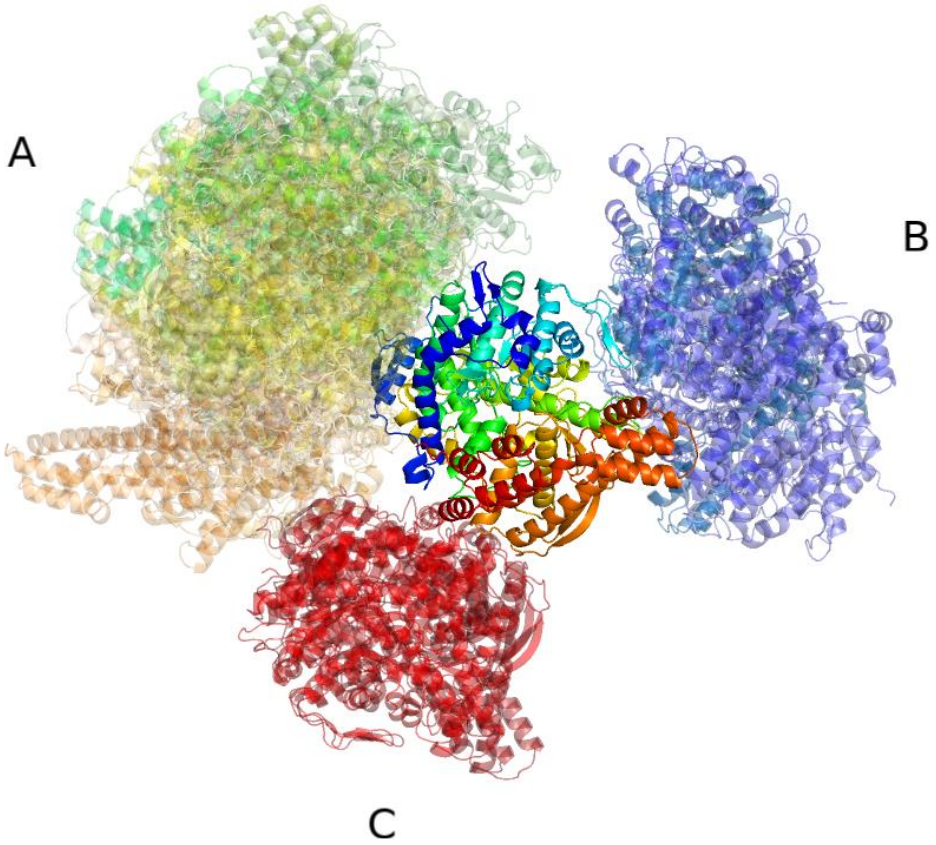

Figure 4. Three surfaces predicted as potential dimerization interfaces. Top 26 predicted dimer arrangements for the RoseTTAFold model using ClusPro are shown, with one subunit of each dimer prediction superimposed (centrally located and colored blue to red from the N- to the C-terminus). The green-yellow-tan cluster of structure shown at the top left comprises most of the predicted structures, and also includes the lowest-energy dimer prediction shown in figure 5. However, two other locations were identified as potential dimerization interfaces and are shown as a blue cluster on the right, or as a red cluster to the lower left of the superimposed subunit.[31]

The lowest-energy predicted dimer is located within the green-yellow-tan cluster of structures shown in Figure 4. The location and orientation of this dimer suggest that the extended N-terminus of *S. acidocaldarius* malate synthase is involved in forming the dimerization interface **(Figure 5)**. Interestingly, this interface also corresponds to the location of intersubunit interactions observed in the only experimentally determined structure of an oligomeric malate synthase reported. This is the malate synthase from *Haloferax*

volcanii, determined by X-ray crystallography, which was shown to form active trimers and hexamers using gel-filtration chromatography.[9] The hexameric form is composed of two back-to-back trimers as observed in the crystal structure. When one of the subunits in the *H. volcanii* hexamer is overlain onto one of the subunits in the predicted dimer shown in Figure 5, the locations of the respective subunit interfaces can be compared directly. The interface in this lowest-energy predicted dimer structure of *S. acidocaldarius* corresponds to a position on the *H. volcanii* subunit that interacts with two different subunits in the opposing trimer within the hexamer. Conversely, the locations of the red and the blue clusters in Figure 4 don't correspond to any of the subunit interfaces within the *H. volcanii* hexameric arrangement. These observations further support the possibility that the predicted dimer shown in Figure 5 may represent the true dimer interface. It is important to note that the presence and location of this dimer arrangement is only a prediction and cannot be confirmed without experimentally solving the protein structure.

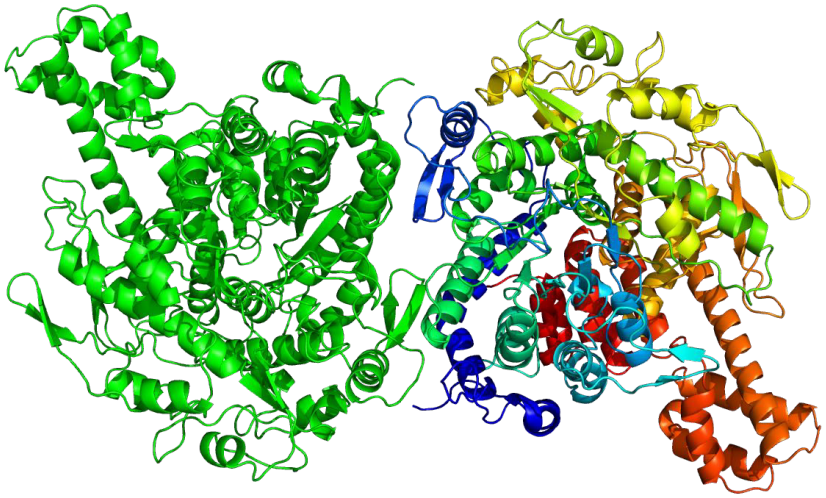

Figure 5. Top predicted RoseTTAFold dimer arrangement, as predicted by ClusPro, viewed along an approximate two-fold rotation axis of symmetry, with one subunit colored blue to red from the N- to the C-terminus. N-terminal segments (shown in dark blue in the monomer on the right) come together to largely form the dimerization interface.[30, 31]

In addition to the RoseTTAFold server, we also utilized the AlphaFold structure prediction. Using PyMOL, we performed an overlay of the structures predicted by each algorithm **(Figure 6)**. When comparing the structures, we noticed differences in the spatial orientation of some secondary structures. While the structures are not identical, they are very similar, including the conserved structure of the TIM barrel forming the active site. Most of the protein is structured in generally the same way in both predictions. Most of these differences between the predictions can be explained by slight shifts and rotational movements of the various portions of the proteins. However, one helical subdomain, shown in the lower right side of **Figure 6**, is rotated differently relative to the main TIM barrel domain.

In conclusion, we found strong evidence for a conserved mechanism in the malate synthase of *S. acidocaldarius*, suggesting that magnesium is required in the active site. The ClusPro server prediction suggests that the extended N-terminus of *S. acidocaldarius* malate synthase is involved in forming a dimerization interface. In addition, we found slight differences between the predicted models from RoseTTAFold and AlphaFold. These findings are based on predicted structures, not confirmed protein structures, and results cannot be confirmed without an experimentally solved protein structure. But based on our detailed comparisons, we anticipate that Arg-214 acts as a bidentate cation to bridge the two substrates and stabilize the oxyanions formed during the catalytic cycle, Asp-600 acts as the catalytic base, and both Asp-326 and Glu-298 chelate the suggested Mg^{2+} cofactor.

The differences found between the RoseTTAFold and AlphaFold predictions underscore the need for an experimentally determined protein structure. This will allow verification of the predicted structure and mechanism involved in the malate synthase of *S.*

acidocaldarius. An experimentally solved structure will confirm the existence of a dimer and the details of the interface. Additionally, this would provide an opportunity to determine which algorithm (RoseTTAFold or AlphaFold) produced the more accurate prediction.

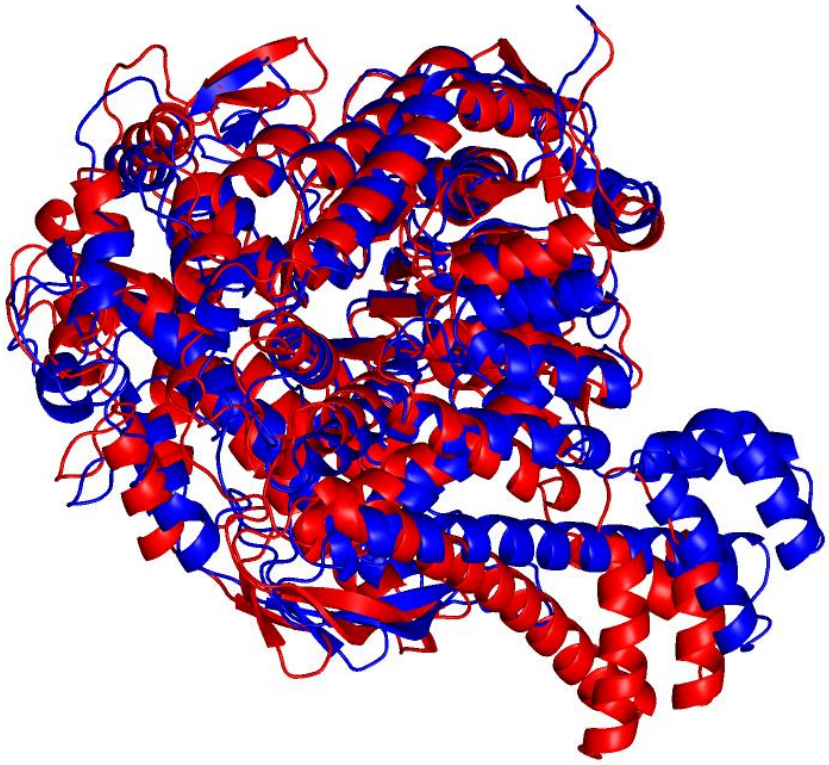

Figure 6. Overlay of the RoseTTAFold monomer model (blue) and the AlphaFold monomer model (red). The overlay was performed by superimposing the C-alpha atoms of four catalytic residues in each monomer as described in the methods section and shown in Figure 3. Both monomers are shown as backbone cartoon traces to clarify the structural variations between these two predicted structures.[31]

ACKNOWLEDGMENTS

The authors thank the Walter Maxwell Gibson Family endowment at SUU for providing a research fellowship to Shaelee Nielsen in support of this work.

REFERENCES

1. Kornberg, H. L., and Krebs, H. A. (1957) Synthesis of cell constituents from C2-units by a modified tricarboxylic acid cycle. *Nature* 179, 988-991. *https://doi.org/10.1038/179988a0*
2. Ajl, S. J. (1956) Conversion of acetate and glyoxylate to malate. *J Am Chem Soc* 78(13), 3230-3231. *https://doi.org/10.1021/ja01594a079*
3. Kornberg, H. L., and Beevers, H. (1957) A mechanism of conversion of fat to carbohydrate in castor beans. *Nature* 180(4575), 35–36. *https://doi.org/10.1038/180035a0*
4. Nakazawa, M., Minami, T., Teramura, K., Kumamoto, S., Hanato, S., Takenaka, S., Ueda, M., Inui, H., Nakano, Y., and Miyatake, K. (2005) Molecular characterization of a bifunctional glyoxylate cycle enzyme, malate synthase/isocitrate lyase, in Euglena gracilis. *Comp Biochem Physiol B Biochem Mol Biol* 141, 445-452. *https://doi.org/10.1016/j.cbpc.2005.05.006*
5. Serrano, J. A., Camacho, M., and Bonete, M. J. (1998) Operation of glyoxylate cycle in halophilic archaea: presence of malate synthase and isocitrate lyase in Haloferax volcanii. *FEBS Lett* 434(1-2), 13-16. *https://doi.org/10.1016/s0014-5793(98)00911-9*

6. Kornberg, H. L. (1966) The role and control of the glyoxylate cycle in Escherichia coli. *Biochem J* 99(1), 1-11. *https://doi.org/10.1042/bj0990001*

7. McKinney, J. D., Höner zu Bentrup, K., Muñoz-Elías, E. J., Miczak, A., Chen, B., Chan, W. T., Swenson, D., Sacchettini, J. C., Jacobs, W. R., Jr., and Russell, D. G. (2000) Persistence of Mycobacterium tuberculosis in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. *Nature* 406(6797), 735–738. *https://doi.org/10.1038/35021074*

8. Lorenz, M. C., and Fink, G. R. (2001) The glyoxylate cycle is required for fungal virulence. *Nature* 412, 83-86. *https://doi.org/10.1038/35083594*

9. Bracken, C. D., Neighbor, A. M., Lamlenn, K. K., Thomas, G. C., Schubert, H. L., Whitby, F. G., and Howard, B. R. (2011) Crystal structures of a halophilic archaeal malate synthase from Haloferax volcanii and comparisons with isoforms A and G. *BMC Struct Biol* 11(23), 1-19. *https://doi.org/10.1186/1472-6807-11-23*

10. Vanderwinkel, E., and De Vlieghere, M. (1968) Physiology and genetics of isocitritase and the malate synthases of Escherichia coli. *Eur J Biochem* 5(1), 81-90. *https://doi.org/10.1111/j.1432-1033.1968.tb00340*

11. Lohman, J. R., Olson, A. C., and Remington, S. J. (2008) Atomic resolution structures of Escherichia coli and Bacillus anthracis malate synthase A: Comparison with isoform G and implications for structure-based drug discovery. *Protein Sci* 17(11), 1935–1945. *https://doi.org/10.1110/ps.036269.108*

12. Durchschlag, H., Biedermann, G., and Eggerer, H. (1981) Large-scale purification and some properties of malate synthase from bakers-yeast. *Eur J Biochem* 114, 255–262.

13. Khan, A.S., Van Driessche, E., Kanarek, L., and Beeckmans, S. (1993) Purification of the glyoxylate cycle enzyme malate synthase from maize (Zea mays L.) and characterization of a proteolytic fragment. *Protein Expr Purif* 4(6), 519-28. *https://doi.org/10.1006/prep.1993.1068*

14. Mullakhanbhai, M.F., and Larsen, H. (1975) Halobacterium volcanii spec. nov., a Dead Sea halobacterium with a moderate salt requirement. *Arch Microbiol* 104, 207–214. *https://doi.org/10.1007/BF00447326*

15. Serrano, J. A., and Bonete, M. J. (2001) Sequencing, phylogenetic and transcriptional analysis of the glyoxylate bypass operon (ace) in the halophilic archaeon Haloferax volcanii. *Biochim Biophys Acta* 1520(2), 154–162. *https://doi.org/10.1016/s0167-4781(01)00263-9*

16. Baliga, N. S., Bonneau, R., Facciotti, M. T., Pan, M., Glusman, G., Deutsch, E. W., Shannon, P., Chiu, Y., Weng, R. S., Gan, R. R., Hung, P., Date, S. V., Marcotte, E., Hood, L., and Ng, W. V. (2004) Genome sequence of Haloarcula marismortui: a halophilic archaeon from the Dead Sea. *Genome Res* 14(11), 2221–2234. *https://doi.org/10.1101/gr.2700304*

17. Thomas, G., Lamlenn, K., and Howard, B. R. (2009) Crystallization and preliminary x-ray diffraction of a halophilic archaeal malate synthase. *AJUR* 8(2 & 3), 15–23.

18. Howard, B. R., Endrizzi, J. A., and Remington, S. J. (2000) Crystal structure of Escherichia coli malate synthase G complexed with magnesium and glyoxylate at 2.0 A resolution: mechanistic implications. *Biochemistry* 39(11), 3156–3168. *https://doi.org/10.1021/bi992519h*

19. Brock, T. D., Brock, K. M., Belly R. T., and Weiss, R. L. (1972) Sulfolobus: A new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Arch Microbiol* 84, 54-68. *https://doi.org/10.1007/BF00408082*

20. Chen, L., Brügger, K., Skovgaard, M., Redder, P., She, Q., Torarinsson, E., Greve, B., Awayez, M., Zibat, A., Klenk, H. P., and Garrett, R. A. (2005) The genome of Sulfolobus acidocaldarius, a model organism of the Crenarchaeota, *J Bacteriol* 187(14), 4992–4999. *https://doi.org/10.1128/JB.187.14.4992-4999.2005*

21. Uhrigshardt, H., Walden, M., John, H., Petersen, A., and Anemüller, S. (2002) Evidence for an operative glyoxylate cycle in the thermoacidophilic crenarchaeon Sulfolobus acidocaldarius. *FEBS Lett* 513(2-3), 223–229. *https://doi.org/10.1016/s0014-5793(02)02317-7*

22. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583-589. *https://doi.org/10.1038/s41586-021-03819-2*

23. Baek, M., DiMaio, F., Anishcenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch L. N., Schaeffer, R. D., Millán, C., Park, H., Adams, C., Glassman, C. R., DeGiovanni, A., Pereira, J. H., Rodrigues A. V., van Dijk, A. A., Ebrecht, A. C., Opperman, D. J., Sagmeister, T., Buhlheller, C., Pavkov-Keller, T., Rathinaswamy, M. K., Dalwadi, U., Yip, C. K., Burke, J. E., Garcia, K. C., Grishin, N. V., Adams, P. D., Read, R. J., and Baker, D. (2021) Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373:6557, 871-876. *https:doi.org/10.1126/science.abj8754*

24. Desta I. T., Porter K. A., Xia B., Kozakov D., and Vajda S. (2020) Performance and its limits in rigid body protein-protein docking. *Structure* 28(9), 1071-1081. *https://doi.org/10.1016/j.str.2020.06.006*

25. Vajda S., Yueh C., Beglov D., Bohnuud T., Mottarella S. E., Xia B., Hall D. R., and Kozakov D. (2017) New additions to the ClusPro server motivated by CAPRI. *Proteins* 85(3), 435-444. *https://doi.org/10.1002/prot.25219*

26. Kozakov D., Hall D. R., Xia B, Porter K. A., Padhorny D., Yueh C., Beglov D., and Vajda S. (2017) The ClusPro web server for protein-protein docking. *Nat Protoc* 12(2), 255-278. *https://doi.org/10.1038/nprot.2016.169*

27. Kozakov D., Beglov D., Bohnuud T., Mottarella S., Xia B, Hall D. R., Vajda, S. (2013) How good is automated protein docking? *Proteins* 81(12), 2159-66. *https://doi.org/10.1002/prot.24403*

28. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Cryst* 66(4), 486-501. *https://doi.org/10.1107/S0907444910007493*

29. Emsley P., and Cowtan K. (2004) Coot: Model-building tools for molecular graphics. *Acta Cryst* 60(Pt 12 Pt 1), 2126–2132.

30. Yueh C., Hall D. R., Xia B., Padhorny D., Kozakov D., and Vajda S. (2017) ClusPro-DC: Dimer classification by the Cluspro server for protein–protein docking. *J Mol Biol* 429(3), 372-381. *https://doi.org/10.1016/j.jmb.2016.10.019*

31. Schrödinger, L., and DeLano, W. (2020) *PyMOL*. Retrieved from *http://www.pymol.org/pymol*

32. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N, Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000) The RCSB Protein Data Bank, www.rcsb.org. Nucleic Acids Research 28, 235–242.

ABOUT THE STUDENT AUTHORS

Jantzen Orton graduated from Southern Utah University in April 2022, earning a Bachelor of Science in Biology and a minor in Chemistry. He is attending medical school at the Texas Tech University Health Sciences Center.

Shaelee Nielsen is currently attending Southern Utah University. She plans to graduate with a Bachelor of Science in Biology in Spring 2024 and then complete medical school.

PRESS SUMMARY

Our project focuses on how a malate synthase enzyme works in a microorganism called *Sulfolobus acidocaldarius*. Most malate synthase enzymes need magnesium to function properly, but one study reported that the malate synthase in this microorganism does not. This is interesting because it suggests a new way the enzyme could work. Recently, new methods have been developed that use artificial intelligence to predict protein structure. We used these new methods to predict the structure of the malate synthase enzyme and found the structure to be very similar to the structures of the other malate synthases. This suggests that the malate synthase in *Sulfolobus acidocaldarius* has a function similar to the other malate synthases, suggesting that magnesium is required for proper functioning.

The Impact of Narratives on Healthcare Decision-Making in Online Discourse

Zayd Almaya & Tom Mould*

Department of History, Anthropology, & Classics, Butler University, Indianapolis, IN

https://doi.org/10.33697/ajur.2024.104

Student: zalmaya@butler.edu
Mentor: tmould@butler.edu

ABSTRACT

This study examines first what type of evidence is most influential in online discussions for patients when making decisions about their health and second how people deploy, interpret, and react to stories in these online discussions to better understand the role and importance of narrative in the medical field. Data was gathered on the platform Reddit using the subreddit r/melanoma for a duration of two weeks. 242 posts were collected and analyzed. Using a combination of grounded theory and coding criteria from sociologist and narrative scholar Francesca Polletta, a code book was developed and applied to all 242 posts to assess narrative impact and engagement. Results demonstrate that evidence based on past experiences and factual information were the most persuasive. Additionally, stories yielded greater discussion, greater empathetic connections, and greater positive responses from online discussants than other forms of evidence. Further, those positive responses indicate that patients seeking medical advice were more likely to express agreement with the advice when it was offered with a story. Given these results, greater attention should be paid to narratives shared in online communities, particularly considering the levels of misinformation and disinformation found online and the evolving relationships between doctors and patients where authority is no longer so easily assumed.

KEYWORDS

Narrative; Personal Experience; Fact; Evidence; Persuasion; Medical Decision-Making; Social Media

INTRODUCTION

The doctor-patient relationship is an essential part of medical care. It is the method by which symptoms are observed and discussed, diagnoses are made, and patient care is offered. Historically, the doctor-patient relationship has followed a paternalistic model in which the physician dominates all aspects of the medical decision-making process. Until recently, it was widely accepted in the medical community that only the physician could truly understand the scope of symptoms and disease and accurately make a diagnosis.[1] This was largely due to the assumption that the general public lacked sufficient medical knowledge. However, within the past few decades, various doctor-patient relationships that emphasize patient participation have become increasingly common. This transition can be traced back to the work conducted by psychologists Josef Breuer and Sigmund Freud that highlighted the significance of the patient as a unique person rather than a group of symptoms. They discussed the importance of communication with the patient as a critical step in diagnosis. This then placed the patient at the center of the doctor-patient relationship and hence led to patient-centered medicine.[1]

In 1956, Szasz and Hollender outlined three distinct models of doctor-patient relationships, legitimizing the view that it was not necessary for physicians to act authoritatively[2]:

- The first model is the activity-passivity model, which refers to traditional paternalism in the medical field. This model involves a power imbalance where the physician acts in the "patients' best interest." The patient has little to no autonomy or say in medical decision-making. This model has roots dating back to Ancient Egypt and for centuries dominated the medical field.[1]

- The second model is the guidance-cooperation model. The physician offers guidance and instruction to the patient due to their extensive medical knowledge. The patient is then expected to comply with the advice offered to them. Similar to the previous model, the patient holds little power in this relationship, yet the physician is not authoritative.

- The third model is the mutual participation model. Both the physician and the patient share responsibility for making medical decisions and planning the treatment. This relationship is characterized by equal power and consideration for the other's values and expectations. This model is seen most commonly in cases of chronic disease or pain in which the patient can maintain greater independence.

The introduction of these psychosocial theories that emphasized a communicative relationship between physicians and patients and subsequently the creation of various doctor-patient models has led to increased patient autonomy and involvement in medical decision-making.

Due to this shift toward increased patient engagement, scholars have begun to consider additional factors in the medical decision-making process when evaluating patient involvement. On the provider side, factors include the issue of empathy and the degree to which physicians respond positively or negatively in response to patients' needs, concerns, and feedback.[3]

There are also patient-related factors that impact medical decision-making including education level and health literacy. Patients having a lower literacy level are more likely to ask the physician to repeat something but "less likely to ask a substantive question, use medical terminology, refer to medications by name, request additional services, or seek new information."[4] There are also demographic factors that impact the extent of patient participation including age and sex as research shows that younger patients and females are more active in the decision-making process.[3]

Important social, cultural, and personal experiences also impact this process, including but not limited to, the patient's socioeconomic status, race, medical preferences, behavior, and concerns.[5] This is particularly the case in terms of patient's perception of their vulnerability to an "unintended or unexpected incident which could have or did lead to harm for the patient."[3] Historically, such incidents have disproportionately impacted racial and ethnic minorities in the U.S.[6,7] In order to reduce the frequency of these incidents, a growing number of patients have demanded a larger role in the decision-making process.[3]

The type of medical decision also impacts the degree to which patients are involved in the decision-making process. Dr. Raisa Deber explains that "problem-solving" and "decision-making" are two distinct areas for patient participation. Deber claims that problem-solving is not suitable for the patient, as there is only one correct answer requiring extensive medical knowledge. In the example of a radiograph that displays a broken leg, she argues, there is only one correct diagnosis: a broken leg.[8] These situations, therefore, are not entirely conducive to patient participation in solving the problem at hand and conducting the diagnosis. However, physicians still utilize the patients' experiences in making the diagnosis and these diagnoses are not always so straightforward or accurate, as one of the most common types of patient narratives in describing their medical history is misdiagnosis stories.[9]

Decision-making situations require the patient to evaluate the value of potential outcomes and their preference for each,[8] presenting a greater opportunity for patient participation than diagnosis situations. One study concerning patient preferences in involvement in medical decision-making reveals that patients had a stronger desire to be involved in decisions that require no medical knowledge.[10]

Outside of the doctor-patient relationship, there are various factors that influence patient participation in the decision-making process, including social media and online communities. There are many identifiable benefits of social media use to patient participation in medical decision-making, the most common of them being enhanced patient empowerment,[11] defined as "the discovery and development of one's inherent capacity to be responsible for one's own life."[12] Patient empowerment can be characterized by patients developing the necessary skills, resources, and information in order to reach their goals in the medical setting. Social media can complement the information shared by physicians, offer patients a unique platform to express their genuine emotions, and provide them with the opportunity for social comparison, or comparing their condition with other patients.[11] Further, online communities provide patients a space to talk about health issues in lay terms, distinct from the often alienating language of institutional medicine. [13] These effects have led to more equal communication between patients and physicians due to patients' increased confidence, health literacy, and increased willingness to participate in the decision-making process.[14]

However, not all of the effects of social media usage are positive, particularly with respect to the reliability of medical information. Patients are increasingly challenging physicians' expertise. Physicians have often responded by reacting negatively to patients' online activity, which in turn has had a disempowering effect on patients.[14] This spiral of distrust has led some patients to change physicians.[11]

Just as attention to the specific media formats used in medical decision-making is important, so too is the mode. Over the past few decades, there has been an explosion of social science research focused on the role of narrative in shaping human behavior, perception, and knowledge.[15–18] In the medical field, physicians utilize stories in a variety of ways: they hear stories from patients, share their patients' stories with other patients as well as with their colleagues, trade stories of their own experiences with their colleagues, and reflect on all of these stories as they consider their own medical practice.[13] In the doctor-patient interaction specifically, storytelling "is the method by which the meaning of the illness and disease are integrated and interpreted by both the physician and the patient." Further, the patients' stories plays an integral role in establishing their viewpoint in the decision-making process.[19]

However, anthropologists, folklorists, and historians have long discussed the need for further emphasis on storytelling in the medical setting.[19–32] For instance, Dr. Arthur Kleinman, a physician-anthropologist, suggests that doctors must pay close attention to their patients' narratives and move past "clinical interrogation."[33] Additionally, Dr. Rita Charon argues that "narrative competence" is essential to the practice of medicine. She defines narrative competence as "the set of skills required to recognize, absorb, interpret, and be moved by the stories one hears or reads."[34] Both scholars argue that physicians could enhance their clinical and emotional skills through these two practices, leading to better patient care.

Despite this call there has been little research into how stories shared outside of the specific doctor-patient dyad impact medical decision-making. A few studies suggest that narratives increase patient adherence and knowledge. Psychological benefits including increased empathy and improved affective forecasting, have also been identified.[35] Other studies demonstrated that narratives had a significant impact in certain conditions. For instance, narratives shared via an audio or video format were more likely to be persuasive compared to a written narrative.[36] Further, narrative research beyond the medical field specifically suggests powerful impacts, not least of which in terms of persuasion. For example, Jones and Crow found that personal beliefs and emotions play an essential role in decision-making and that decisions are not solely based on scientific evidence. In terms of persuasiveness, therefore, they argue that communicators would benefit by presenting information in a narrative format.[37]

The increasing prevalence of online communities and chat forums dedicated to particular medical issues and illnesses have created a space for people to discuss their health issues, including seeking advice for treatment. Narratives shared in these spaces have been shown to lead to improvements in health care initiatives for patients,[38] provide emotional support,[39,40] increase personal empowerment,[41] provide a positive model for socializing new online group members,[40] and alternatively promote health literacy[39,40] as well as risk perpetuating false information.[42]

Based on the current scholarly literature, it is clear that analysis of the role of narrative in medical decision-making is needed, particularly in new media. Accordingly, this study examines the relationship between narratives shared on social media and patient decision-making, asking the following two questions:

1 - What types of evidence shared on social media are persuasive when making decisions about one's healthcare?

2 - What role do stories play in discussions of healthcare on social media?

For the purposes of this study, we define narrative as a text that indexes or describes a past event; provides a sequence of action; includes elements of orientation, complication, and resolution; and is structured with a beginning, middle, and end. Narratives include plots, characters, settings, and actions. Structured as bounded events, narratives can be retold in new settings, making them particularly easy to share.[43–47] This definition adheres to the relatively strict definitions used in folklore studies, linguistic anthropology, and sociolinguistics that attend closely to form. Looser definitions used increasingly often throughout the social sciences may not include these qualities, and as such, may not share the same outcomes of more strictly defined narratives.[48]

METHODS AND PROCEDURES

The study utilized a mixed-methods approach in order to describe the relationship between various types of evidence used in online discussions and how the patient reacted to the advice that was offered.

There were many social media platforms and diseases that could have been selected and investigated for this research. Ultimately, however, examining the subreddit r/melanoma on the Reddit platform was selected for three reasons. First, its discussion-based algorithm provides an opportunity for multiple types of discourse and facilitates conversations among users about virtually any topic including medical care.[49] Second, the nature of the discussion and questions posed were frequently geared towards identification and diagnosis, both of which were directly or indirectly related to medical decision-making. (Many of the other social media sites dedicated to medical issues focused on lifestyle and relationships post-diagnosis, which does not involve medical decision-making.) Third, the subreddit r/melanoma was both a large and active group, with approximately 5,500 members and multiple new posts daily. Accordingly, this social media site offered an accessible, relevant, and sizable data set.

Coding began on September 25, 2022, working with posts beginning two full weeks earlier to help avoid missing newly posted comments. Coding included all posts between September 9, and September 15, 2022, reaching saturation point when no new patterns or types of posts emerged, and after which a significant body of data had been collected.

There were 41 original posts, with 201 responses, resulting in 242 total posts in the dataset. Of the 242 posts, 67 contained a story: 19 shared in the original post, 48 in the comments. However, not all 48 stories in the comments were shared by audience members: 17 stories were shared by the original poster (OP), 31 by commenters.

After the dataset was compiled, open coding techniques were used to identify emergent themes and patterns. Additionally, coding categories from Francesca Polletta's work assessing narrative persuasiveness in political discourse were included in order to compare this data in the context of her findings. Eventually, a formal codebook was constructed that included the number of upvotes (likes), number of comments, inclusion of an image, type of post (question or comment), type of question, whether the post included a story, function of the story, tone, nature of response, whether

advice was offered, types of evidence, types of authority, and, most importantly, the original poster's response. To ensure intercoder reliability, both co-authors independently coded the first 38 posts (17% of the dataset) and compared results. Of the 13 coding categories across 38 posts, there was disagreement on only 17 out of the 507 total number of codes. At 97% alignment, this is well above the 90% typically needed for reliability. The codebook was then clarified to address the disagreement and then applied consistently to the entire dataset. Statistical analyses were then applied in order to look for additional patterns, correlations, and frequencies.

RESULTS

The purpose of this study was to evaluate the significance of various types of evidence utilized in online communities when patients are making decisions about their health as well as to specifically assess the nature and impact of narratives. While the data made it impossible to determine the medical decision the patient ultimately made, it was possible to determine how they reacted (positively, negatively, or no response at all), and in some cases what they *said* they were going to do.

Types of Evidence

The most prevalent evidence types utilized in the online posts were observational evidence (e.g. "It definitely changed shape, and it doesn't look like any other mole I have") (30.2%), past experience-based evidence (e.g. "I'm a diagnosed hypochondriac and I'm telling you this, melanoma just doesn't often look anything like what you've got pictured here") (16.9%), and opinion-based evidence (e.g. "Ah, well I would just get it checked anyway, just to be safe") (15.3%) (see Table 1). The least common evidence types were longitudinal ("It appeared around 2 months ago it has grown I think a little in size") (8.7%), factual (e.g. "CDKN2A mutation increase lifetime risk for getting melanoma") (8.3%), and weblinks (e.g. "https://notamole.com/uglyduckling") (3.7%).

Evidence Types	All Posts	Commenters Only	Advice Shared
Past Experience	16.9% (41)	23.7% (31)	31.0% (18)
Fact	8.3% (20)	11.5% (15)	15.5& (9)
Observational	30.2% (73)	22.9% (30)	32.8% (19)
Longitudinal	8.7% (21)	3.8% (5)	-
Weblink	3.7% (9)	6.1% (8)	13.8% (8)
Opinion	15.3% (37)	28.2% (37)	58.6% (34)

Table 1. Frequency of Evidence Type Table displays percentages and (counts), Sample sizes are as follows: Dataset = 242, Commenters = 131, Advice Shared = 58

Table 1 displays how often each evidence type was utilized across all of the sampled posts, including original posters (column 1). The frequency of evidence types utilized solely by commenters was examined as well in order to more accurately determine what evidence types are being utilized in response to original posters. Opinion-based (28.2%), past experience-based (23.7%), and observational evidence (22.9%) were the three most common evidence types that commenters used. This was followed by factual (11.5%), weblink-based (6.1%), and longitudinal evidence (3.8%) (column 2). Finally, the frequency of evidence types utilized by commenters specifically when sharing advice to the original posters was determined as well (column 3). This displays that opinion-based evidence was by far the most prevalent (58.6%), followed by observational evidence (32.8%), past experience-based evidence (31.0%), factual (15.5%), and weblink-based evidence (13.8%).

More significantly, the OP positive response rate for each evidence type when advice was shared was also calculated in an effort to determine the most persuasive evidence (Table 2). In this study, OP positive response rate was utilized in order to gauge the original poster's reception to various advice shared in response to their question. Some examples of responses that indicated a positive response by the original poster include:

Evidence Type	Positive OP Response Rate
Observational	32.8%
Past Exeperience	50.0%
Fact	44.4%
Weblink	12.5%
Opinion	32.4%

Table 2. Positive OP Response Rate per Evidence Type

"SUCH a helpful article! You my friend are the reason I'm putting away my phone right now and going to sleep."

"Alright, I'll see if I can get an appointment! Thank you for the replies."

The data show that past experience-based evidence had the most favorable response rate (50.0%), with factual evidence (44.4%) as a close second. This was followed by observational evidence (32.8%) and opinion-based evidence (32.4%). Lastly, weblink-based evidence had the lowest response rate (12.5%). There was only one instance of longitudinal evidence being utilized when advice was being offered to the OP, therefore, this metric was not included when evaluating positive OP response rate.

Types of Evidence & Narrative

The study's research questions focused on types of evidence and narrative. Therefore, narrative status and prevalence of evidence usage are important factors in determining persuasiveness (Table 3).

Advice Type and Frequency	Positive OP Response Rate
Non-narrative	24.4%
Narrative	53.8%*
Single Evidence	25.0%
Multiple Evidence	38.5%

Table 3. Positive OP Response Rate when Advice is Shared *Significance was reported for the following level: $*p < 0.05$ [$t(56) = 2.1$, $p = 0.04$]

Virtually all of the stories shared were personal experience narratives: first-person stories shared by the person involved (95.5%; the other 4.5% were secondhand stories from family and close friends). When stories were shared there was a positive OP response rate of 53.8%, compared to a rate of 24.4% when stories were not shared. Narrative advice, in comparison to non-narrative advice, received statistically significantly higher positive responses from the original poster. On average, multiple forms of evidence when offering advice received a positive OP response rate of 38.5%. However, when only a single type of evidence was used, this rate dropped to 25% (Table 3), suggesting, not surprisingly, that more evidence is better than less.

Further, analysis of these two factors combined makes it clear how narrative and multiple types of evidence work together to impact positive OP response rate (Table 4).

Narrative and Evidence Status	Positive OP Response Rate
Non-narrative & Single evidence	23.1%
Non-narrative & Multiple evidence	26.3%
Narrative & Single evidence	33.3%
Narrative & Multiple evidence	71.4%

Table 4. Positive OP Response Rate based on Storytelling and Number of Evidence Types when Advice is Shared

When advice was offered, yet no story was shared and only one type of evidence was utilized, there was a positive

OP response rate of 23.1%. When no story was shared and multiple types of evidence were used, the rate increased to 26.3%. The same trajectory was apparent when stories were shared, but the increase with multiple evidence was much more dramatic. When there was a story shared with one type of evidence, this rate was 33.3%. However, when multiple types of evidence were used, and a story was shared, this more than doubled the positive OP response rate to 71.4%.

Narrative Usage

The following data focus solely on the second research question related to narrative use: when and why are stories shared, and how are stories received. Stories were most often shared in response to explicit diagnosis questions, followed by implicit diagnosis questions, and then treatment questions (Table 5). However, relative to the number of questions that were shared for each question type, questions related to treatments yielded the most stories: on average, 4 stories were shared per treatment question.

Question Type	Number of Stories	Stories to Question Ratio
Implicit Diagnosis	15	1.07
Explicit Diagnosis	21	0.88
Treatment	12	4

Table 5. Frequency of Stories

In terms of function, narratives were used to inform audiences, answer questions, offer advice, offer empathy, express anxiety, validate concerns, build credibility, persuade readers, and offer examples (Table 6). The vast majority of the stories (97.0%) were used to inform. However, these functions are not mutually exclusive, and often, stories would contain multiple functions.

Function of Narratives	Frequnecy
Inform	97.0%
Answer Question	32.8%
Offer Advice	17.9%
Empathetic	9.0 %
Express Anxiety	16.4%
Share Similar Concern	14.9%
Build Credibility	7.5%
Persuade	4.5%
Offer Example	9.0%

Table 6. Function of Narratives

In terms of reception by the original poster, the 31 stories shared by commenters received a positive OP response rate of 61.3% compared to the positive response rate of the entire dataset of 44%. Further, although 61.3% of responses were positive, this does not mean the remaining responses were negative. In fact, there were no negative responses to stories that were shared by commenters. Rather, the remaining responses were either neutral (32.3%) or there was no response (9.7%).

The dataset was also analyzed according to functions identified by Francesca Polletta in her study of narratives used in online discourse. Analysis included the responses among all participants (not just the original poster) as well as non-narrative claims in order to accurately assess the impact that narratives have on their audiences. The 67 stories in the dataset yielded 71 responses from audience members.

This data show that the majority of the functions were more prevalent in response to narrative compared to non-narrative

Function	Percentage of Narrative Claims	Percentage of Non-Narrative Claims
Acknowledge Emotional Impact	12.7%**	3.5%
Make a Corroborating Claim	11.2%	5.8%
Share a Similar Concern	1.4%	4.7%
Answers Question	35.2%	36.6%
Ask a Follow-up Question	32.4%***	9.9%
Express Appreciation	9.9%	5.2%
Disagree	1.4%	2.3%

Table 7. Audience Members' Reception of Narratives *The percentages for each response type were calculated by dividing the number of each response by the number of claims in its category. There were 71 narrative claims and 172 non-narrative claims. **Significant on the 0.01 level, $t(241) = 2.8$, $p = .006$ ***Significant on the 0.0001 level, $t(241) = 4.5$, $p = .00001$

claims. In response to narrative claims, the most common response was answering OP's question (35.2%), followed by asking a follow-up question (32.4%), acknowledging emotional impact (12.7%), making a corroborating claim (11.2%), and expressing appreciation (9.9%). Lastly, disagreeing, asking to clarify, and sharing the same concern had the same frequency (1.1%). Particularly relevant to this study, posts in response to narrative claims were significantly more likely to acknowledge the emotional impact and to ask a follow-up question in comparison to posts in response to non-narrative claims.

The 172 responses to non-narrative posts were also analyzed. This revealed that the most common response was answering OP's question (36.6%), then asking a follow-up question (9.9%). This was followed by making a corroborating claim (5.8%), sharing the same concern (4.7%), expressing appreciation (5.2%), acknowledging emotional impact (3.5%), disagreement (2.3%), and asking to clarify (0.5%).

DISCUSSION

There were two main objectives of this study: to understand 1) which evidence types utilized in online communities are most frequent and most persuasive in medical decision making, and 2) the role narratives in particular play in these contexts. From the beginning, assessing persuasiveness was difficult. It was not possible to determine whether people actually followed the advice given, even when offering positive feedback such as appreciation for their advice or recognition of their peace of mind. However, this study does offer insight into the types of evidence used most often, the degree to which that evidence prompted a positive response, and the impact narrative had in these online communities compared to other types of evidence.

In terms of the prevalence of evidence types in the dataset, observational (30.2%) and past experience-based evidence (16.9%) were the most common, both of which are commonly used with narratives. Even in posts that did not include a narrative, these types of evidence far surpassed factual evidence (8.3%) and statistical evidence, of which there was not a single instance. Further, none of those posting commented on the low degree or complete lack of facts or statistics, suggesting that evidence that is primarily anecdotal and personal - both of which lend themselves to narrative - is the expectation for these online communities. The reliance on individual-based evidence (personal past-experience and opinion) is even more clear when distinguishing between the commenters and the overall dataset, revealing that opinion-based evidence (28.2%) is the most prevalent, followed by past experience (23.7%) and observational evidence (22.9%). When focusing solely on posts that offered advice to the original poster, the pattern is even stronger, with opinion-based evidence as the most prevalent evidence type (58.6%), followed by observational (33.0%), and past experience (31.0%). Although narratives were rarely used among posts that were offering advice (22.0%), evidence types common with narratives were more common than factual evidence.

In terms of the types of evidence that resulted in a positive response rate from the original poster, however, past experience-based evidence remained high (0.5), with factual evidence a close second (0.444). This reveals that these two evidence

types were the two most impactful and significant for patients. Although factual evidence was not common when advice was shared (0.155), it was responded to positively by posters (44.4%). Furthermore, it is understandable that observational evidence resulted in a low positive response (0.328) rate as it was used broadly and frequently in the dataset. The same can be said about opinion-based evidence (0.324), as there was no experience or factual basis to a commenter's advice. Yet, it is interesting to note that weblinks (any link to a website with medical information) contained the lowest positive response rate (0.125). It would be expected that they would be similarly received as factual evidence, if not in a more positive manner due to the potential of their increased credibility, yet this was not the case. This may suggest an unwillingness to follow external links in these contexts. Additionally, there is no relationship between prevalence and persuasiveness, as the most common evidence type utilized when advice was shared (opinion-based) was not the most persuasive.

Due to the frequency and high positive response rate of evidence types common with stories, further relationships among posts that were offering advice to the original poster were explored, first looking at the correlation between positive response rates and storytelling. This revealed that narrative advice in the aggregate contained a significantly higher positive response rate compared to non-narrative advice (Table 3), although stories are not shared as frequently, they are far more effective in persuading people in online discourse. Non-narrative advice was more than three times more prevalent than narrative advice (3.5:1), yet the positive response rate for non-narrative advice (0.244) was less than half that of narrative advice (0.538), signifying the importance of stories in this setting.

One possible explanation for this dramatic finding is the function of stories as "empathy machines." Stories invite audiences to view a situation from another person's vantage point, encouraging an empathetic response. Previous research has shown that empathy plays a vital role in persuading audience members of one's argument generally, and in the medical decision-making process specifically. [49] This interpretation is supported by the significantly higher positive response rate for narrative advice in comparison to non-narrative advice.

Shifting to consider the amount rather than the type of evidence types reveals not surprisingly that more evidence is better than less in terms of positive response rate. While this was true for non-narrative posts as well, the increase from 33.3% to 71.3% was dramatic and significant in establishing not only the impact of narrative, but narrative's reliance on additional forms of evidence for effective persuasion (Table 4). This information, alongside the significantly large positive OP response rate, leads us to assert the persuasiveness and effectiveness of storytelling when making decisions about one's health.

In terms of when stories were shared, treatment-related questions yielded the highest number of stories (4), although they were by far the least prevalent question type. (Table 5). This could be explained by the fact that diagnosis questions tend to be original poster-focused, whereas treatment questions are "you" focused and invite audience members to speak about their own experiences, including by sharing a narrative. Stories allow us to be egocentric and talk about ourselves. For example, a typical treatment-related question was, "Any members in here ever had atypical moles removed?" This question invited audience members to share their own experience with atypical mole removal to answer the original poster's question, hence the much higher story-to-question ratio. Conversely, an example of a diagnosis question was, "Should I be worried? This mole is quite new and has gotten bigger over the past couple of weeks." This question is specific to the original poster and therefore yields less personal sharing by audience members and subsequently fewer stories.

In order to understand why stories were shared, each post was coded for its function. The variety of functions displayed in Table 6 demonstrates the breadth of narrative functions in online communities. The function 'to inform' was the most common as original posters informed their audience members of their condition as they asked their question, and audience members informed the original poster and other readers of their past history in their response.

Another common function was to 'express anxiety,' also commonly used by original posters concerned for their health.

Further, this function was most common in diagnosis questions. In fact, all but 2 narratives that expressed anxiety were from diagnosis questions. Furthermore, it is significant that the 'offer exemplar' and 'empathetic' functions contained high positive OP response rates (83.3%). Both of these functions are frequently found in narrative.

In terms of how people seeking medical advice responded to what people had to say, stories resulted in a notably larger positive response rate (61.3%) compared to the overall dataset (44.3%). This suggests that the use of narratives may increase the chance that advice is responded to positively.

In terms of understanding how audience members as opposed solely to the original poster reacted and responded to stories, Polletta's model of assessing narrative function and responses was utilized and applied to this study. Nearly every function was more prevalent in response to narrative claims compared to non-narratives (Table 8). Additionally, posts in response to narrative claims were significantly more likely to acknowledge the emotional impact of the narratives and include follow-up questions compared to non-narrative posts, highlighting the importance of narratives in increasing collaboration and discourse in online medical communities.

Function	Narrative Claims	Narrative Claims*	Non-Narrative Claims	Non-Narrative Claims*
Acknowledge Emotional Impact	12.7%	6.0%	3.5%	0.5%
Make a Corroborating Claim	11.2%	18.1%	5.8%	1.4%
Express Appreciation	9.9%	26.4%	5.2%	16.9%
Disagree	1.4%	14.3%	2.3%	5.1%

Table 8. Comparison of the Rates of Reception of Narrative and Non-Narrative Claims *Data From Polletta (2011)

In order to compare these findings to other research concerning narrative impact, this analysis was compared to Polletta's research and a similar pattern from this study can be observed. Acknowledging emotional impact, making a corroborating claim, and expressing appreciation were all more prevalent in response to stories in both studies. The sample size utilized in this study was smaller and the general response rate that had been calculated as a part of this study is generally lower compared to Polletta's work, yet increased conversations and collaborations among members in response to stories can be observed.

Comparison also suggested some differences that may be related to the differences between political and medical discourse. In this study, the frequency of posts that acknowledged the emotional impact was greater in comparison to Polletta's work, which could be the case due to the stress and anxiety-inducing nature of medical discourse. Yet, in the political discussions in Polletta's work, making a corroborating claim and disagreement were both more common, perhaps due to the natural argumentative nature of political conversations. Additionally, there was a significant difference in the frequency of posts that expressed appreciation, with rates much higher in Polletta's work. This difference may be attributed to the fact that the Reddit platform in this study allows people to upvote (like) posts. More than 95% of posts received at least one upvote, for both narrative and non-narrative claims. Such a finding is not particularly useful, but it does offer a possible explanation for why the numbers in this study were so much lower than in Polletta's since an upvote could serve as a form of expressing appreciation, replacing the need to explicitly state it in one's comment.

CONCLUSIONS

With the growing prevalence and popularity of online communities and chat forums dedicated to various medical conditions and the significant impact that narratives have been shown to have on persuasive discourse, storytelling on social media outside of the doctor-patient dyad needs to be considered and evaluated when discussing patient involvement and participation in the medical decision-making process. This study suggests that although narrative advice is not as prevalent as non-narrative advice, its persuasive impact is significantly greater. Further, narratives encourage greater interaction by acknowledging emotional impact, making corroborating claims, and expressing appreciation. These findings suggest that if people want others to heed their advice, they should share stories, ideally with multiple forms of ev-

idence. If doctors want to establish collaborative and trusting relationships with their patients, they will likely need to engage in storytelling as well in order to compete with the stories their patients are hearing online. In either case, understanding the medical decision-making process in the U.S. today will require greater attention to how stories shared online shape our perceptions and therefore our behaviors.

REFERENCES

1. Kaba, R., Sooriakumaran, P. (2007) The evolution of the doctor-patient relationship, *International Journal of Surgery* 5, 57–65. *https://doi.org/10.1016/j.ijsu.2006.01.005*
2. Szasz, T., Hollender, M. (1956) A Contribution to the Philosophy of Medicine; the Basic Models of the Doctor-Patient Relationship, *AMA Arch Intern Med* 97, 585–592. *https://doi.org/10.1001/archinte.1956.00250230079008*
3. Davis, R., Jacklin, R., Sevdalis, N., Vincent, C. (2007) Patient Involvement in Patient Safety: What Factors Influence Patient Participation and Engagement?, *Health Expect* 10, 259–267. *https://doi.org/10.1111/j.1369-7625.2007.00450.*
4. Katz, M., Jacobson, T., Veledar, E., Kripalani, S. (2007) Patient Literacy and Question-Asking Behavior During the Medical Encounter: a Mixed-Methods Analysis, *J Gen Intern Med* 22, 782–786. *https://doi.org/10.1007/s11606-007-0184-6*
5. Hajjaj, F., Salek, M., Basra, M., Finlay, A. (2010) Non-Clinical Influences on Clinical Decision-Making: a Major Challenge to Evidence-Based Practice, *J R Soc Med* 103, 178–187. *https://doi.org/10.1258/jrsm.2010.100104*
6. Baciu, A., Negussie, Y., Geller, A., Weinstein J. (2017) The State of Health Disparities in the United States, in *Communities in Action: Pathways to Health Equity*, National Academies Press
7. Lister, R., Drake, W., Scott, B., Graves, C. (2019) Black Maternal Mortality-The Elephant in the Room, *World J Gynecol Womens Health* 3. *https://doi.org/10.33552/wjgwh.2019.03.000555*
8. Deber, R. (1994) The Patient-Physician Partnership: Decision Making, Problem Solving and the Desire to Participate, in *Physicians in Health Care Management*, 423–427, Canadian Medical Association Journal
9. Mulley, A., Trimble, C., Elwyn, G. (2012) Stop the Silent Misdiagnosis: Patients's Preferences Matter, *BMJ* 345:e6572. *https://doi.org/10.1136/bmj.e6572*
10. Thompson, S., Pitts, J., Schwankovsky, L. (1993) Preferences for Involvement in Medical Decision-Making: Situational and Demographic Influences, *Patient Education and Counseling* 22, 133–140. *https://doi.org/10.1016/0738-3991(93)90093-c*
11. Smailhodzic, E., Hooijsma, W., Boonstra, A., Langley, D. (2016) Social Media Use in Healthcare: A systematic Review of Effects on Patients and on Their Relationship with Healthcare Professionals, *BMC Health Services Research* 16 *https://doi.org/10.1186/s12913-016-1691-0*
12. Wentzer, H., Bygholm, A. (2013) Narratives of Empowerment and Compliance: Studies of Communication in Online Patient Support Groups, *International Journal of Medical Informatics* 82, 386-394 *https://doi.org/10.1016/j.ijmedinf.2013.01.008*
13. Goldstein, D. (2000) When Ovaries retire: Contrasting women's experiences with feminist and medical models of menopause, *Health* 4, 309-323
14. Steiner, J. (2005) The use of Stories in Clinical Research and Health Policy, *JAMA* 294, 2901–2904. *https://org.10.1001/jama.294.22.2901*
15. Czarniawska, B. (2004) The Uses of Narrative in Social Science Research, in *Handbook of Data Analysis*, 649–666, SAGE Publications.
16. Hurwitz, B., Greenhalgh, T., Skultans, V. (2008) The Temporal Construction of Medical Narratives, in *Narrative Research in Health and Illness*, 414–427, Blackwell Publications
17. Lai, C. (2010) Narrative and Narrative Enquiry in Health and Social Sciences, *Nurse Researcher* 17, 72–84. *https://doi.org/10.7748/nr2010.04.17.3.72.c7748*
18. Overcash, J. (2003) Narrative Research: A Review of Methodology and Relevance to Clinical Practice, *Critical Reviews in Oncology/Hematology* 48, 179–184. *https://doi.org/10.1016/j.critrevonc.2003.04.006*

19. Roter, D. (2000) The Enduring and Evolving Nature of the Patient-Physician Relationship, *Patient Education and Counseling* 39, 5–15. *https://doi.org/10.1016/s0738-3991(99)00086-5*

20. Blank, T., Kitta, A. (2015) *Diagnosing Folklore: Perspectives on Disability, Health, and Trauma*, University Press of Mississippi.

21. Bock, S. (2013) Staying Positive: Women's Illness Narratives and the Stigmatized Vernacular, *Culture and Society* 5, 150 – 166.

22. Charon, R. (2008) *Narrative Medicine: Honoring the Stories of Illness*, Oxford University Press

23. Eggly, S. (2002) Physician-Patient Co-Construction of Illness Narratives in the Medical Interview, *Health Communication* 3, 339 – 360.

24. Goldstein, D. (2001) Competing Logics and the Construction of Risk, in *Healing Logics: Culture and Medicine in Modern Health Belief Systems*, Utah State University Press

25. Goldstein, D. (2015) Vernacular Turns: Narrative, Local Knowledge, and the Changed Context of Folklore, *Journal of American Folklore* 508, 125 – 145.

26. Gray, J. (2009) The Power of Storytelling: Using Narrative in the Healthcare Context, *Journal of Communication in Healthcare* 3, 258 – 273.

27. Hovey, R., Paul, J. (2007) Healing, the Patient Narrative-Story and the Medical Practitioner: A Relationship to Enhance Care for the Chronically Ill Patient, *International Journal of Qualitative Methods* 4, 53 – 66.

28. Hyden, L., Bulow, P. (2006) Medical Discourse, Illness Narratives, *Encyclopedia of Language and Linguistic* 2, 697 – 703.

29. Kitta , A. (2011) *Vaccinations and Public Concern in History: Legend, Rumor, and Risk Perception*, Routledge Studies in the History of Science, Technology, and Medicine

30. Lucius-Hoene, G., Thiele, U., Breuning, M., Haug, S. (2012), Doctors' Voices in Patients' Narratives: Coping with Emotions in Storytelling, *Chronic Illness* 3, 163 – 175.

31. Johanna, S. (2011) Illness Narratives: Reliability, Authenticity, and the Empathic Witness, *Medical Humanities* 2, 68 – 72.

32. Sharf, B., Vanderford, M. (2003). Illness Narratives and the Social Construction of Health, *Handbook of Health Communication*, 9 – 34.

33. Tanner, D. (1999). The Narrative Imperative: Stories in Medicine, Illness and Bioethics, *HEC* 2, 155 – 169.

34. Kleinman, A. (1997) Medical Anthropology as Intellectual Career, in *Writing at the Margin: Discourse between Anthropology and Medicine*, University of California Press.

35. Charon, R. (2004) Narrative and Medicine, in *New England Journal of Medicine* 350, 862–864. *https://doi.org/10.1056/NEJMp038249*

36. Shaffer, V., Brodney, S., Gavaruzzi, T., Zisman-Ilani, Y., Munro, S., Smith, S., Thomas, E., Valentine, K., Bekker, H. (2021) Do Personal Stories Make Patient Decision Aids More Effective? An Update from the International Patient Decision Aids Standards, in *Medical Decision Making* 41, 897–906. *https://doi.org/10.1177/0272989X211011100*

37. Shen, F., Sheer, V., Li, R. (2015) Impact of Narratives on Persuasion in Health Communication: A Meta-Analysis, in *Journal of Advertising* 44, 105-113. *https://doi.org/10.1080/00913367.2015.1018467*

38. Jones, M., Crow, D. (2017) How can we use the 'Science of Stories' to Produce Persuasive Scientific Stories?, in *Palgrave Communications*, 3, 53. *https://doi.org/10.1057/s41599-017-0047-7*

39. Sundstorm, B. (2016) Voices of the "99 Percent": The Role of Online Narrative to Improve Health Car, *The Permanente Journal* 4, 15 – 224.

40. Drewniak, D., Glassel, A., Hodel, M., Biller-Andorno, N. (2020) Risks and Benefits of Web-Based Patient Narratives: Systematic Review, *Journal of Medical Internet Research* 3.

41. Kessler, S., Schmidt-Weitmann, S. (2021) Diseases and Emotions: An Automated Content Analysis of Health Narratives in Inquiries to an Online Health Consultation Service, *Health Communication* 2, 226 – 235.

42. Hamilton, H. (2002) Reported Speech and Survivor Identity in On-Line Bone Marrow Transplantation Narratives, *Journal of Sociolinguistics* 2, 53 – 67.

43. Polletta F., (2206) Stories and Reasons: Why Deliberation is Only Sometimes Democratic, in*It was Like a Fever: Storytelling in Protest and Politics*, 98 – 101, the University of Chicago Press.
44. Kitta, A., Goldberg, D. (2016) The Significance of Folklore for Vaccine Policy: Discarding the Deficit Model, *Critical Public Health* 27, 507 – 514.
45. Mieke, B. (1985) *Narratology: Introduction to the Theory of Narrative*, University of Toronto Press.
46. Bruner, J. (1986) *Actual Minds, Possible Words*, Harvard University Press.
47. Freytag, G. (1863) *Freytag's Technique of the drama: An Exposition of Dramatic Composition and Art*, Johnson Reprint Corporation.
48. Labov, W., Waletzky, J. (2015) Narrative Analysis: Oral Versions of Personal Experience, *Journal of Narrative and Life History* 7, 3 – 38.
49. Mould, T. (2020) *Overthrowing the Queen: Telling Stories of Welfare in America*, Indiana University Press.
50. O'Donnell, N., Guidry, J. (2020) BeTheMatch: Assessing How Testimonials on Reddit Promote the Importance of Donating Bone Marrow, *Journal of Health Communication* 25, 660–670. *https://doi.org/10.1080/10810730.2020.1836088*

ABOUT THE STUDENT AUTHOR

Zayd Almaya is a junior undergraduate biology major with Spanish and anthropology minors in the College of Liberal Arts and Sciences at Butler University.

PRESS SUMMARY

Within the last few decades, the doctor-patient relationship has evolved tremendously and has begun to include a greater role for the patient. Additionally, online spaces discussing various diseases have increased in popularity for many patients. Therefore, based on research in medical anthropology, this study examined what type of evidence - medical-based or narrative-based - is most influential for patients making decisions about their health in online discussions. Further, the study applies narrative theories developed by rhetoricians, political scientists, and folklorists to consider how people deploy, interpret, and react to stories in these online discussions to better understand the role and importance of narratives in the medical field. Results demonstrate that past experience and factual-based evidence were the most persuasive, yet stories yielded greater discussion among members, greater empathetic connections among members, and most significantly, increased the chance that medical advice was taken. Therefore, greater attention should be paid to narratives shared in online communities.

Elongation Factor P is Required for Processes Associated with *Acinetobacter* Pathogenesis

*Dylan Kostrevski & Anne Witzky**

Department of Mathematics, Computer, and Natural Sciences, Ohio Dominican University, Columbus, OH

https://doi.org/10.33697/ajur.2024.105

Student: smithd25@ohiodominican.edu
*Mentor: witzkya@ohiodominican.edu**

ABSTRACT
Antibiotic resistance is one of the world's fastest-growing and most prevalent problems today. The influx of antibiotics within our environment from inadequate antibiotic stewardship has led to a surge of drug-resistant microorganisms. The CDC has classified Carbapenem-resistant *Acinetobacter* (CRA) as an urgent threat within this crisis. New drug development is imperative to combat infections caused by drug-resistant pathogens such as CRA. Bacterial translation, the process of protein synthesis by the ribosome, is a common target for new antibiotic development. Elongation factor P (EF-P) is a universally conserved translation factor required for antibiotic resistance in many bacteria. In this study, we assessed the importance of EF-P in processes associated with *Acinetobacter* pathogenesis. In the absence of EF-P, *Acinetobacter baylyi* displays decreased biofilm formation, surface-associated motility, and resistance to beta-lactams and carbapenems. This data holds hope for future drug development targeting EF-P in pathogens closely related to *A. baylyi*.

KEYWORDS
Acinetobacter baylyi; Translation; Ribosome; Elongation Factor P; Polyproline; Biofilm; Surface Associated Motility; Antibiotic Resistance

INTRODUCTION

Antibiotic resistance is one of the world's most pressing problems today because of the difficulty of treating antibiotic-resistant infections. Antibiotic resistance refers to the ability of bacteria to develop resistance to particular drugs designed to kill them.[1] Thus, resistant bacteria will survive in the presence of current antibiotic treatments.[2] The growth of antibiotic-resistant bacteria can be attributed to antibiotic overuse and poor stewardship during hospital discharge.[3] The Center for Disease Control and Prevention (CDC) estimates a national total of 2,868,700 infections and 35,900 deaths each year from resistant fungi and bacteria.[2] The inability to effectively treat patients and the number of deaths display the importance of finding new ways to treat antibiotic-resistant bacteria.

The CDC characterizes bacteria on a threat level ranging from concerning, serious, and urgent based on factors including clinical impact, availability of effective antibiotics, transmissibility, and incidence.[2] Carbapenem-resistant *Acinetobacter* (CRA) bacteria jumped from a threat level of serious to an urgent threat level between the two most recent antibiotic resistance threat reports.[2,4] CRA cause bloodstream infections and pneumonia in patients with weakened immune systems, such as long-term hospital patients in intensive care units.[5] CRA microbes are often multidrug-resistant, meaning the bacteria are resistant to more than just the carbapenem class of antibiotics.[2] These resistant bacteria were responsible for nearly 8,500 hospitalizations and 700 deaths in 2017 according to the CDC. [2]

The increase in infections from multi-drug resistant microorganisms limits the treatments available to a patient.[3] New drug development is imperative to cope with the rise of multi-resistant microbes.[6] Many new drugs in development target translation.[7] Translation is the cellular process that allows both eukaryotic and prokaryotic cells to synthesize proteins using ribosomes.[8-9] Bacterial translation is an intriguing target for adjuvant therapies that could potentiate existing drugs, as bacteria are often hypersensitive to antibiotics in the absence of key a translation factor known as Elongation Factor P (EF-P).[10-13] Elongation factor P (EF-P) is a universally conserved protein that alleviates ribosomal pausing at polyproline motifs.[14-15] Polyproline motifs cause ribosomal pausing because steric hindrance of proline delays the formation of peptide bonds.[16] The strength of the pause is determined by the translation initiation rate, the surrounding amino acids, and the location of the pause on the transcript.[17-19] When the ribosome pauses at a polyproline motif, EF-P enters the vacant E site of the ribosome and aids in the peptide bond formation to allow translation to continue.[20]

In many organisms, EF-P is required for a wide range of cellular processes including growth, motility, and antibiotic resistance.[10-13, 21-22] Here, we characterize the role of EF-P in *Acinetobacter baylyi.* While it is generally non-pathogenic, *A. baylyi* is an established genetic model system that shares a significant genetic similarity with CRA species *Acinetobacter baumanii.*[23] *A. baylyi* is often used in drug resistance studies, and 80% of *A. baumannii* core genes have orthologs in *A. baylyi.*[23] Studying *A. baylyi* could lead to fundamental discoveries applicable to *A. baumannii* and the creation of therapeutics targeting EF-P in CRA. The shared genetic material in combination with being an ideal model organism makes *A. baylyi* an alternative way to study the dangerous opportunistic pathogen *A. baumannii.* In *A. baylyi,* loss of EF-P results in a decreased growth rate, but other phenotypic effects have not been characterized.[24] Here, the physiological significance of EF-P in *A. baylyi* was investigated, focusing on processes that are required for pathogenesis in *Acinetobacter* species.[25-26]

MATERIALS AND PROCEDURES

Bioinformatics

The EF-P protein sequences for *A. baylyi* (Q6FAA9) and *A. baumannii* (B2HVL6) were accessed through UniProt.[27] Alignments of the EF-P protein sequences were generated by a multiple sequence alignment tool on Clustal Omega, and the percent identity between the two sequences was noted.[28] Alignment was repeated comparing *A. baylyi* to the following microorganisms: *Bacillus subtills* (P49778)*, Escherichia coli* (P0A6N4)*, Pseudomonas aeruginosa* (Q9HZZ2)*, Salmonella typhimurium* (P64036)*, Erwinia amylovora* (D4I1J7)*, Neisseria meningitidis* (P0DUK0)*,* and *Staphylococcus aureus* (Q6G937). All percent identities relative to *A. baylyi* were noted.

The full proteomes for *A. baylyi* (UP000000430) and the other microorganisms of interest were accessed through UniProt. Proteins containing PPX motifs were identified using the Expasy ScanProsite tool.[29] The percentage of proteome containing PPX motifs was calculated by dividing the number of proteins on the PPX list by the total number of proteins indicated by UniProt. The percentage PPX was noted for all the following organisms: *A. baumannii* (UP000005740)*, B. subtilis* (UP000001570)*, E. coli* (UP000000558)*, P. aeruginosa* (UP000002438)*, N. meningitidis* (UP000000425), *S. typhimurium* (UP000001014)*, S. aureus* (UP000008816), and *E. amylovora* (UP000001841).

Strains and Growth Conditions

Both the wild-type and *Δefp* mutant strains of *A. baylyi* were graciously donated by Dr. Valrie de Crecy-Lagard in the Microbiology department at the University of Florida.[24] The wild-type strain was ADP1 and the experimental strain was PS6433 (ADP1 *Δefp::sacBKanR*). Both bacterial cultures grew in Lennox formulation Luria broth (LB) (Carolina) liquid tubes with shaking. The wild-type culture grew at 30 °C in a liquid LB culture for 24 hours. The *Δefp* culture grew at 30 °C in liquid LB in the presence of 15 μg/mL kanamycin (Thermo Scientific) for 48 hours. All assays were conducted at 30 °C except for the surface-associated motility assay, which was conducted at 37 °C.

Surface Associated Motility

The motility testing was performed with semisolid agar plates containing 0.5% granulated tryptone, 0.25% NaCl, and 0.3% agarose.[25] A needle was used to inoculate the agar plates by puncturing the top surface of the agar. The bacteria grew in a dense halo formation with jagged edges displaying surface-associated motility. The plates grew in an incubator at 37 °C and were scanned after 24 and 48 hours. The assay was performed in triplicate using biological replication.

Biofilm Formation

Saturated starting cultures for both strains were diluted to an OD of 0.5. 30 μl of each dilution was added to a corresponding three mL LB liquid test tube and placed in the stationary incubator at 30 °C for 72 hours. The liquid media was removed, and adherent cells were stained for 10 minutes in three mL of 0.02% crystal violet dye. The stain was removed, and photographs were taken of each test tube's biofilm formations. One mL of PBS was added to each formation and vortexed for five minutes to collect adherent cells. The cell suspension was collected and the OD_{580} of each sample was measured. The assay was performed in triplicate using biological replication.

Disc Diffusion

Saturated wild-type and *Δefp* cultures were back diluted to an OD_{600} of 0.1. Bacterial lawns of *A. baylyi* were grown on LB plates using sterilized swabs for each sample. Paper discs saturated with 10μg of the following antibiotics were placed on the LB agar plates using forceps: ampicillin (Biogram), amoxicillin (Sensi-Disc), meropenem (Sensi-Disc), and imipenem (Sensi-Disc). The plates were incubated at 30 °C for 48 hours. Plates were scanned and the zones of inhibition for each antibiotic were measured using ImageJ software.[30] The assay was performed in triplicate using biological replication.

RESULTS

EF-P sequence and overall PPX content is similar between A. baylyi and A. baumannii

To support the notion that EF-P likely plays a similar role in both *A. baylyi* and *A. baumannii*, we began by analyzing both the EF-P protein sequence and the polyproline content of both organisms. EF-P in *A. baylyi* shares 92% sequence identity to *A. baumannii* (Table 1). *A. baylyi* and *A. baumanni* also had nearly identical polyproline content in their respective proteomes (26.5%) (Table 1). Taken together, this supports the idea that EF-P will likely have a similar role in both organisms.

To predict the physiological significance of EF-P in *A. baylyi,* we compared the EF-P sequence and polyproline content to organisms in which EF-P has been thoroughly characterized. To function efficiently, EF-P requires post-translational modification at a conserved residue. However, this modification varies widely between different organisms.[31] Organisms that employ *R*-β-lysine modification strategy shared the highest sequence identity and polyproline content with *A. baylyi* (*E. coli, S. typhimurium, E. amylovora*) (Table 1). In contrast, organisms that use *L*-rhamnose modification of EF-P maintained less sequence identity and substantially higher polyproline content in their proteomes (*N. meningitidis* and *P. aeruginosa*) (Table 1). Organisms that use 5-aminopentanol modification had relatively lower percent identity and polyproline content (*B. subtilis* and *S. aureus*) (Table 1). Taken together, these results suggest that in *A. baylyi* EF-P will have a similar functional and physiological role as it does in other organisms that also modify EF-P with the *R-β*-lysine modification.

Table 1. EF-P amino acid percent identity and percentage of proteome containing at least one PPX motif.

Organism	Percent Identity	Percent Proteome with PPX
A. baylyi	100	26.5
A. baumannii	92	26.7
E. coli	60	32.7
S. typhimurium	60	33.6
E. amylovora	59	31.3
B. subtilis	43	21.4
S. aureus	40	16.0
P. aeruginosa	35	45.0
N. meningitidis	31	30.8

EF-P is required for surface associated motility

To demonstrate whether EF-P is utilized in surface motility of *A. baylyi*, we collected data on wildtype and ⊿*efp* strain motility at 24 hours of incubation. Qualitative imaging at 24 hours displayed normal wildtype motility over time, while ⊿*efp* motility was completely abolished (**Figure 1a**). Quantitative measurements of the zones indicated that the difference in motility between wildtype and Δ*efp* were significant (**Figure 1b**). Since the surface associated motility assay is sensitive to differences in growth, we continued to allow all biological replicates to grow until a 48-hour timepoint. If the apparent ⊿*efp* motility defect was a result of decreased growth rather than decreased motility, we would expect the cells to eventually migrate from the inoculation point given sufficient growth time. For two of the ⊿*efp* biological replicates, no surface associated motility was observed; however, one ⊿*efp* replicate displayed a low number of cells that migrated from the inoculation point, included in **Figure 1 (labeled suppressor)**. Given the distinct morphology of the replicate, we believe this was a suppressor mutation to circumnavigate the EF-P requirement for surface associated motility.

EF-P is necessary for successful biofilm formation

To assess the role of EF-P in the biofilm formation, both wildtype and Δ*efp* strains were grown in cultures to promote biofilm formation. After three days of incubation, cells were stained and observed. The WT image indicates a faint stained ring of biofilm formation halfway up the image, whereas in ⊿*efp,* there is no visible indication of biofilm formation (**Figure 2a**). To quantitatively assess this phenotype, stained cells were harvested and the OD_{580} was recorded in triplicate. **Figure 2b** indicates that wildtype strain produced greater biofilm mass than the ⊿*efp* strain. It is possible that the decrease in biofilm mass in the ⊿*efp* strain could be the indirect result in the growth rate defect rather than a direct biofilm formation defect. While we cannot definitely exclude this possibility, it is important to note that there was no detectable biofilm biomass on the tube (absorbance readings were comparable to background). Given the time scale of the experiment, we would expect some level of biofilm formation to have occurred within this timeframe if the mutant strain was able to form a biofilm.

EF-P is required for antibiotic resistance

To determine the role EF-P plays in antibiotic resistance, we analyzed the antibiotic sensitivity of *A. baylyi* to beta-lactams and carbapenems. Each zone of inhibition was larger in *Δefp,* showing more sensitivity to each of the antibiotics than the wildtype control (**Figure 3**). The zones of inhibition for both classes of antibiotics indicate the *A. baylyi* cells lacking *efp* are significantly more sensitive to antibiotics.

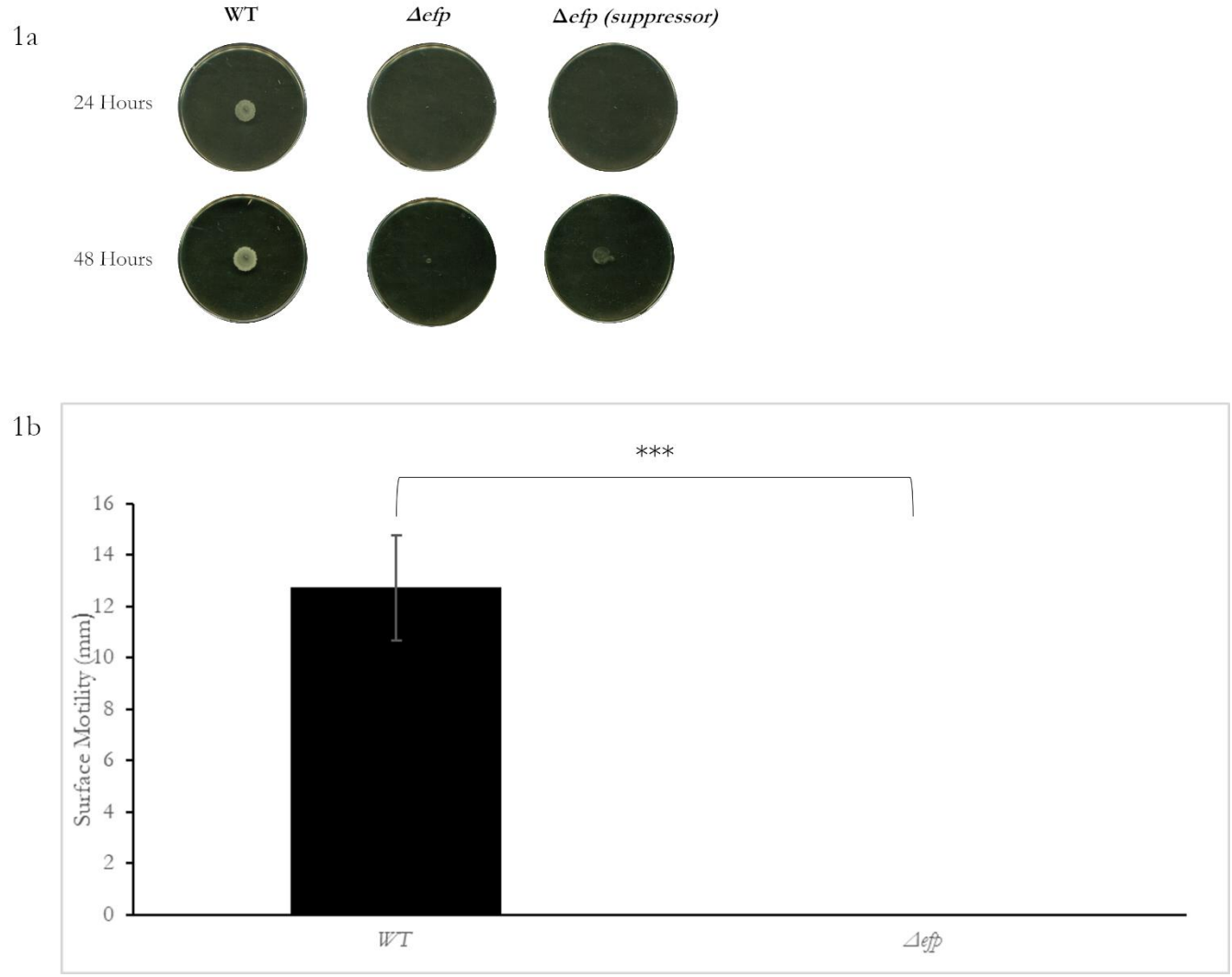

Figure 1. EF-P is required for surface associated motility. (a) WT and Δ*efp* cultures were inoculated in surface associated motility plates in triplicate. Motility was documented after 24 and 48 hours of incubation at 30° C. One replicate produced an atypical migration pattern only after 48 hours, suggesting it acquired suppressor mutation. (b) Quantitative data showing the zone of motility present in the WT and Δ*efp* after 24 hours of incubation. ImageJ was utilized to quantify the surface motility shown in both groups. Since the mutant lacked motility completely, no measurement was collected. A t-test analysis determined there is significance in the data, ***p<0.001. Data represents the mean of three biological replicates. The error bars represent the standard deviation between all three biological replicates.

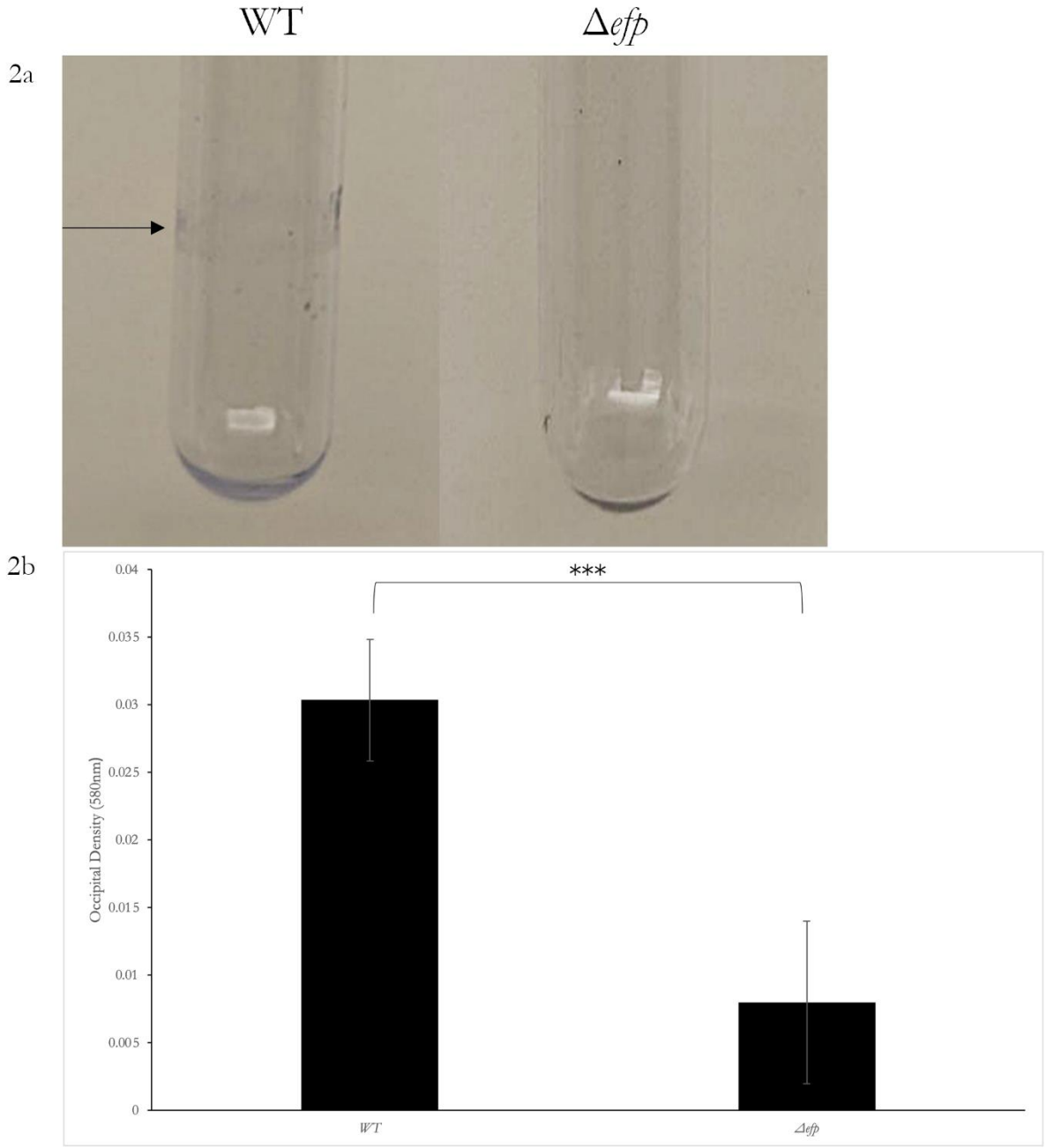

Figure 2. EF-P is required for biofilm formation. (a) Qualitative representation of biofilm ring formation of both wildtype and *Δefp* strains. The photographs were taken after 72 hours of incubation directly followed by 0.02% crystal violet staining. The *Δefp* replicate displays no stained formation, whereas wildtype shows a stained ring biofilm formation as indicated by the arrow. Similar results were obtained for all three replicates. (b) Optical densities at 580 nm shows wildtype biofilm has significantly greater biofilm mass compared to *Δefp*. After resuspending the biofilm cultures with 1 mL of PBS, absorbance readings at an OD of 580 nm were collected. The absorbance readings for *Δefp* were severely lower than wildtype. A t-test analysis determined there is significance in the data, ***$p<0.001$. Data represents the mean of three biological replicates. The error bars represent the standard deviation between all three biological replicates.

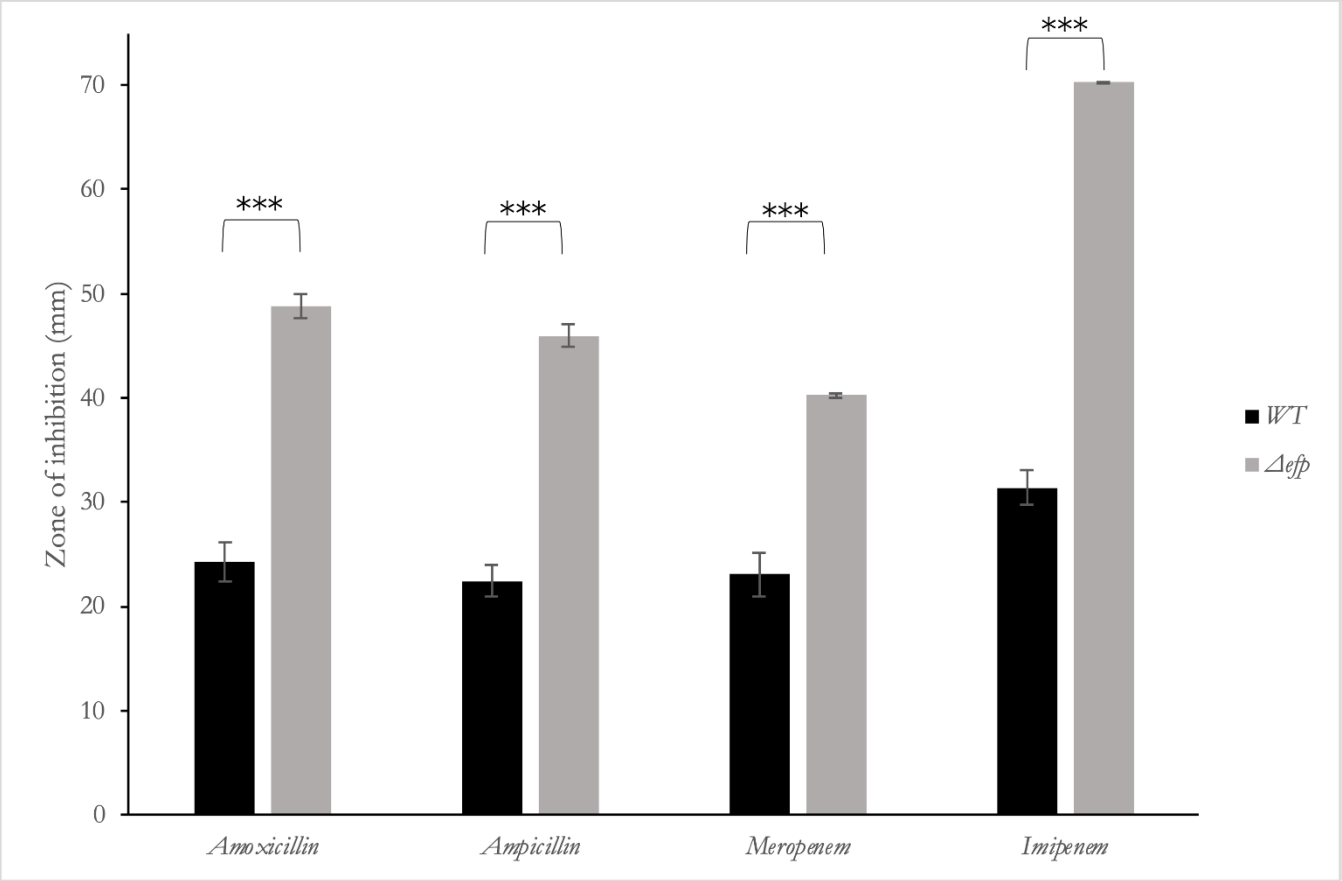

Figure 3. *A. baylyi* is hypersensitive to antibiotics without EF-P. *A. baylyi* zones of inhibition (mm) for amoxicillin, ampicillin, meropenem, and imipenem on both wildtype and Δ*efp* are shown. After growing a bacterial lawn of the strains in the presence of each antibiotic, ImageJ was used to quantify the zones of inhibition for all the samples. The zones of inhibition of each antibiotic are significantly larger in the Δ*efp* group showing less resistance to each antibiotic. A t-test analysis determined there is significance in the data, ***p<0.001. Data represents the mean of three biological replicates. The error bars represent the standard deviation between all three biological replicates.

DISCUSSION

CRA has been identified by the CDC as an urgent threat within the antibiotic resistance crisis.[2] Here, we aimed to characterize the role of EF-P, a potential drug target, in *Acinetobacter* species. Though it is rarely pathogenic, we characterized EF-P in *A. baylyi,* as it has been proposed to serve as a safer model system in studying *A. baumanii.*[23] The percent identity in the EF-P amino acid sequences between both *A. baumannii* and *A. baylyi* is 92%, suggesting EF-P likely functions similarly in both these microorganisms when compared to the other microorganisms surveyed (**Table 1**). After surveying multiple microorganism's proteomes containing a PPX motif, *A. baumannii* and *A. baylyi* contained nearly identical polyproline content within the proteome, displaying even more similarities between the two microorganisms and the likely importance of EF-P in each of these species. (**Table 1**).

In *A. baylyi,* surface associated motility is completely abolished in the absence of EF-P (**Figure 1**). The putative suppressor suggests a biological pathway the bacteria mutated to circumnavigate the requirement of the *efp* gene to successfully perform surface motility. Similarly, biofilm formation and antibiotic resistance are severely impacted by the absence of EF-P (**Figure 2** & **3**). Surface associated motility and biofilm formation are both processes known to be required for pathogenesis in *Acinetobacter* species.[25-26] Antibiotic resistance is of particular interest in a clinical setting considering the prevalence of CRA.[2] Taken together, these phenotypes suggest EF-P plays an important role in the ability of *Acinetobacter* to survive in a clinical setting.

One interesting finding is the disparity between the PPX content of *A. baylyi* and *P. aeruginosa.* Both species have severe growth defects without EF-P, but maintain disparate PPX content in their respective proteomes (**Table 1**).[10] Although one might assume the higher PPX percentage in the proteome is a causation for a greater need of EF-P in a microorganism, we observe *A. baylyi* contains a lower PPX percentage than *P. aeruginosa* but displays equally severe phenotypes without EF-P. However, it is important to also consider the context dependent nature of ribosomal pausing at PPX motifs in the absence of EF-P. The translation

initiation rate, distance from the N-terminus, and the amino acid in the third position of the motif all play a significant role in the severity of ribosomal pausing.[17-19] It is possible that these three factors are contributing to pausing in a high number of proteins in *A. baylyi*. Future ribosomal profiling studies will be required to accurately characterize pausing in the absence of EF-P in *A. baylyi.*

Previous literature has documented phenotypes associated with loss of EF-P within *A. baumannii*.[32] The biofilm results presented here concur with their findings that EF-P is required for biofilm formation. However, it was also established here that in *A. baylyi,* EF-P is also required for surface associated motility. According to Guo et al., the deletion of *efp* enhanced the surface associated motility of *A. baumannii*.[32] There are several possible explanations for this discrepancy. One explanation would be the length of time the surface associated motility tests we conducted. Guo et al. conducted the assay for 8 hours, while here data was collected after 24 and 48 hours. It is possible that the requirement for EF-P varies throughout the stages of this physiological state. A second explanation could be due to the nature of the strain construction in both studies. Here, a true knockout deletion strain of *A. baylyi* was used, while Guo et al. used a deletion that placed *efp* under the control of an inducible promoter. Inducible systems can be subject to leaky expression, allowing for some expression of the target protein.[33] Differences in EF-P expression in these two systems could lead to different results. Lastly, though *A. baylyi* and *A. baumanii* are highly similar, they are distinct species of *Acinetobacter*. Guo et al., has shown that in *A. baumannii,* EF-P interacts with c-di-GMP. However, this interaction does not appear to occur in other microbes. It is plausible that this may also be the case for *A. baylyi,* meaning EF-P is playing a unique role in *A. baumanii*.[32]

CONCLUSIONS

EF-P is required for biofilm formation, surface associated motility, and resistance to beta-lactams and carbapenems in *A. baylyi.* Bioinformatic analyses suggest that EF-P likely maintains a similar role in the closely related pathogen *A. baumanii.* This work shows that EF-P contributes to a wide range cellular processes related to pathogenesis, which may be applicable to *A. baumanii*.[25-26] Development of antibiotics targeting EF-P would likely have minimal off target effects due to a very minimal structural resemblance of eukaryotic EF-P (EIF5A).[34] However, we do not yet understand how exactly to target EF-P, nor do we understand the most important PPX motifs alleviated by EF-P in *A. baylyi*. Future ribosomal profiling studies must be conducted to better understand how each PPX motif impacts the severity of ribosomal pausing in *A. baylyi*.

ACKNOWLEDGEMENTS

The authors thank Dr. Valérie de Crécy-Lagard at the University of Florida for providing strains used in the study.

REFERENCES

1. Munita, J. M. and Arias, C. A. (2016) Mechanisms of antibiotic resistance, *Microbiol Spectr* 4(2):VMBF-0016-2015, *https://doi.org/10.1128/microbiolspec.VMBF-0016-2015*
2. Antibiotic resistant threats in the United States 2019, Centers for Disease Control and Prevention, *https://www.cdc.gov/drugresistance/biggest-threats.html* (Accessed Feb 2023)
3. Vaughn, V. M., Hersh, A. L. and Spivak, E. S. (2022) Antibiotic overuse and stewardship at hospital discharge: the reducing overuse of antibiotics at discharge home framework, *Clin Infect Dis* 74, 1696–1702. *https://doi.org/10.1093/cid/ciab842*
4. Antibiotic resistant threats in the United States 2013, Centers for Disease and Control Prevention, *https://www.cdc.gov/drugresistance/biggest-threats.html* (Accessed Feb 2023)
5. Antunes, L. C. S., Visca, P., and Towner, K. J. (2014) Acinetobacter baumannii: evolution of a global pathogen, *Pathog Dis* 71, 292–301. *https://doi.org/10.1111/2049-632X.12125*
6. Uddin, T. M., Chakraborty, A. J., Khusro, A., Zidan, B. R. M., Mitra, S., Emran, T. B., Dhama, K., Ripon, K. H. Md., Gajdács, M., Sahibzada, M. U. K., Hossain, J. Md., and Koirala, N. (2021) Antibiotic resistance in microbes: History, mechanisms, therapeutic strategies and future prospects, *J Infect Public Health* 14, 1750–1766. *https://doi.org/10.1016/j.jiph.2021.10.020*
7. Wilson, D. (2014) Ribosome-targeting antibiotics and mechanisms of bacterial resistance, *Nat Rev Microbiol* 12, 35–48. *https://doi.org/10.1038/nrmicro3155*
8. Witzky, A., Tollerson, R., and Ibba, M. (2019) Translational control of antibiotic resistance, *Open Biol* 26, 190051. *https://doi.org/10.1098/rsob.190051*
9. Wilson, D. N. and Cate, J. H. D. (2012) The Structure and function of the eukaryotic ribosome, *Cold Spring Harb Perspect Biol* 4, 1–18. *https://doi.org/10.1101/cshperspect.a011536*
10. Rajkovic, A., Erickson, S., Witzky, A., Branson, O., Seo, J., Gafken, P. R., Frietas, M. A., Whitelegge, J. P., Faull, K. F., Navarre, W., Darwin, A. J., and Ibba, M. (2015) Cyclic rhamnosylated elongation factor P establishes antibiotic resistance in *Pseudomonas aeruginosa mBio* 6, e00823-15. *https://doi.org/10.1128/mbio.00823-15*
11. Witzky, A., Hummels, K.R., Tollerson, R., Rajkovic, A., Jones, L.A., Kearns, D.B., and Ibba, M. (2018) EF-P posttranslational modification has variable impact on polyproline translation in *Bacillus subtilis mBio* 3, e00306-e00317. *https://doi.org/10.1128/mbio.00306-18*

12. Tollerson, R., Witzky, A., and Ibba, M. (2018) Elongation factor P is required to maintain proteome homeostasis at high growth rate, *Proc Natl Acad Sci USA* 115, 11072-11077. *https://doi.org/10.1073/pnas.1812025115*

13. Zou, S.B., Hersch, S.J., Wiggers, J.B., Leung, A.S., Buranyi, S., Xie, L.J., Dare, K., Ibba, M., and Navarre, W.W. (2012) Loss of elongation factor P disrupts bacterial outer membrane integrity, *J Bacteriol* 194, 413-425. *https://doi.org/10.1128/jb.05864-11*

14. Ude, S., Lassak, J., Starosta, A., Kraxenberger, T., Wilson, D. N., and Jung, K. (2013) Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches, *Science* 339, 82–85. *https://doi.org/10.1126/science.1228985*

15. Doerfel, L., Wohlgemuth, I., Kothe, C., Peske, F., Urlaub, H., and Rodnina, M. V. (2013) EF-P is essential for rapid synthesis of proteins containing consecutive proline residues, *Science* 339, 85–88. *https://doi.org/10.1126/science.1229017*

16. Doerfel, L., Wohlgemuth, I., Kubyshkin, V., Starosta, A., Wilson, D., Budisa, N., and Rodnina, M.V. (2015) Entropic contribution of elongation factor P to proline positioning at the catalytic center of the ribosome, *J Am Chem* 137, 12997-13006. *https://doi.org/10.1021/jacs.5b07427*

17. Woolstenhulme, C. J., Guydosh, N. R., Green, R. and Buskirk, A. R. (2015) High-precision analysis of translational pausing by ribosome profiling in bacteria lacking EFP, *Cell Rep* 11, 13–21. *https://doi.org/10.1016/j.celrep.2015.03.014*

18. Elgamal, S., Katz, A., Hersch, S.J., Newsom D., White, P., Navarre, W.W., and Ibba, M. (2014) EF-P dependent pauses integrate proximal and distal signals during translation, *Plos Genet* 10, e1004551. *https://doi.org/10.1371/journal.pgen.1004553*

19. Hersch, S.J., Elgamal, S., Katz A., Ibba, M., and Navarre, W.W. (2014) Translation initiation rate determines the impact of ribosome stalling on bacterial protein synthesis, *J Biol Chem* 289, 28160-28171. *https://doi.org/10.1074/jbc.M114.593277*

20. Blaha, G., Stanley, R. E., and Steitz, T. A. (2009) Formation of the First Peptide Bond: The Structure of EF-P Bound to the 70S Ribosome, *Science* 325, 966–970. *https://doi.org/10.1126/science.1175800*

21. Rajkovic, A., Hummels K.R., Witzky, A., Erickson, S., Gafken, P.R., Whitelegge, J.P., Faull, K.F., Kearns, D.B., and Ibba, M. (2016) Translational control of swarming proficiency in *Bacillus subtilis* by 5-amino-pentanolylated elongation factor P, *J Biol Chem* 291, 10976-10985. *https://doi.org/10.1074/jbc.M115.712091*

22. Yanagisawa, T., Takahashi, H., Suzuki, T., Masuda, A., Dohmae, N., and Yokoyama, S. (2016) *Neisseria meningitidis* translation elongation factor P and its active-site arginine residue are essential for cell viability. *PLoS One* 11, e0147907. *https://doi.org/10.1371/journal.pone.0147907*

23. de Berardinis, V., Durot, M., Weissenbach, J., and Salanoubat, M. (2009) Acinetobacter baylyi ADP1 as a model for metabolic system biology, *Curr Opin in Microbiol* 12, 568–576. *https://doi.org/10.1016/j.mib.2009.07.005*

24. de Crécy, E., Matzgar, D., Allen, C., Penicaud, M., Lyons, B., Hansen, C. J., and de Crécy-Lagard, V. (2007) Development of a novel continuous culture device for experimental evolution of bacterial populations, *Appl Microbiol and Biotechnol* 77, 489–496. *https://doi.org/10.1007/s00253-007-1168-5*

25. Corral, J., Perez-Varela, M., Sanchez-Osuna, M., Cortes, P., Barbe, J., and Aranda, J. (2021) Importance of twitching and surface-associated motility in the virulence of *Acinetobacter baumannii*, *Virulence* 12, 2201–2213. *https://doi.org/10.1080/21505594.2021.1950268*

26. Mea, J.H., Yong, P.V.C., and Wong, E.H. (2021) An overview of *Acinetobacter baumannii* pathogenesis: motility, adherence, and biofilm formation, *Microbiol Res* 247, 126722. *https://doi.org/10.1016/j.micres.2021.126722*

27. The UniProt Consortium. (2023) UniProt: the universal protein knowledgebase in 2023, *Nucleic Acids Res* 51, D523-D531. *https://doi.org/10.1093/nar/gkac1052*

28. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, Weizhong, L., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J.D., and Higgins, D.G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, *Mol Syst Biol* 7, 539. *https://doi.org/10.1038/msb.2011.75*

29. De Castro, E., Sigrist, C.J.A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P.S., Gasteiger, E., Bairoch, A., and Hulo, N. (2006) ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins, *Nucleic Acids Res* 34, W362-365. *https://doi.org/10.1093/nar/gkl124*

30. Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012) NIH Image to ImageJ: 25 years of image analysis, *Nat Methods* 9, 671–675. *https://doi.org/10.1038/nmeth.2089*

31. Hummels, K.R. and Kearns, D.B. (2020) Translation elongation factor P (EF-P), *FEMS Microbiol Rev* 1, 208-218. *https://doi.org/10.1093/femsre/fuaa003*

32. Guo, Q., Cui, B., Wang, M., Li, X., Tan, H., Song, S., Zhou, J., Zhang, L. H., and Deng, Y. (2022) Elongation factor P modulates Acinetobacter baumannii physiology and virulence as a cyclic dimeric guanosine monophosphate effector, *Proc Natl Acad Sci USA* 119, e2209838119. *https://doi.org/10.1073/pnas.2209838119*

33. Zhang, Y., Shang, X., Lai, S., Zhang, G., Liang, Y., and Wen, T. (2012) Development and Application of an Arabinose-Inducible Expression System by Facilitating Inducer Uptake in Corynebacterium glutamicum, *Appl Environ Microbiol* 78, 5831–5838. *https://doi.org/10.1128/aem.01147-12*

34. Tillery, L. M., Barrett, K. F., Dranow, D. M., Craig, J., Shek, R., Chun, I., Barrett, L. K., Phan, I. Q., Subramanian, S., Abendroth, J., Lorimer, D. D., Edwards, T. E., and Van Voorhis, W. C. (2020) Toward a structome of *Acinetobacter baumannii* drug targets, *Protein Sci* 29, 789–802. *https://doi.org/10.1002/pro.3826*

ABOUT STUDENT AUTHOR

Dylan Kostrevski is a recent graduate from Ohio Dominican University and achieved his Bachelor of Science (B.S.) in Biology. He plans on taking a gap year to work in industry before continuing his education. After gaining technical experience working, he aims to gain admission to a doctorate program in Molecular Biology from an R01 university.

PRESS SUMMARY

Antibiotic resistance is one of society's most prevalent problems today. According to the Centers for Disease Control and Prevention (CDC), 35,900 people die each year from antibiotic-resistant infections. The CDC ranks Carbapenem-Resistant Acinetobacter (CRA) species as an urgent threat to global health. This study characterizes the physiological significance of a translation factor in *Acinetobacter* species to assess the potential for this protein to be a target in future drug development against CRA.

Measurement System for Compliance in Tubular Structures

Ave Kludze, Anthony R. D'Amato, & Yadong Wang*

Department of Biomedical Engineering, Cornell University, Ithaca, NY

https://doi.org/10.33697/ajur.2024.106

*Students: akk86@cornell.edu**
Mentors: ard223@cornell.edu, yw839@cornell.edu

ABSTRACT
Tubular structures such as blood vessels, intestines, and the trachea are common in various life forms. This paper describes a measurement system to test the mechanical compliance of tubular structures. The novelty of the system lies in its hardware and software. Here we use vascular graft as an example to demonstrate the utility of the system. A fully synthetic vascular graft would ideally mimic the mechanical and architectural properties of a native blood vessel. Therefore, mechanical testing of the graft material under physiological pressure is crucial to characterizing its potential in vivo performance. The device operates through a low-cost Arduino-based control system that simulates and measures cyclic fluid pressure changes over time and a laser micrometer that measures diameter changes with pressure. This system is low-cost, assuming one already has access to a laser micrometer. In contrast to previous methods, this system offers a simple, low-cost, and customizable option to measure compliance and is equipped with data acquisition/analysis programs. These programs include a MATLAB application that processes and synchronizes Arduino Uno pressure signals and LabChart Pro diameter readings. Lastly, this paper explains the hardware and software of the measurement system. The system is beneficial for testing the pressure-diameter relationship of tubular structures of varying sizes and materials.

KEYWORDS
Tubular Structures; Compliance; Data Acquisition System; Physiological Pressure; Diameter Change; Arduino Uno; LabChart Pro; MATLAB

INTRODUCTION

In many lifeforms, tubular structures are essential for organ function, cellular transport, and structural stability.[1] Each tubular structure can perform varying roles based on its location and composition.[1] For instance, the trachea in the respiratory system is comprised of pseudo-stratified columnar epithelium which helps transport and filter air, the esophagus in the digestive system consists of stratified squamous cells that aid in moving foods and liquids, the intestine with its brush border cells in the digestive system absorb nutrients, the ureter in the urinary system composed of urothelium passes urine, and blood vessels in the circulatory system are multilayered to withstand blood pressure and to transport blood.[1]

Recent efforts in tissue engineering have focused on developing tubular structures that can replace damaged organs after accidents or illnesses.[1] Difficulty arises in designing tubular structures that can mechanically mimic their native counterparts and work in environments that are dependent on liquid, solid, or air material exchange.[1] Compliance is an important mechanical property often used to assess the structure-function relationship of tubular vessels. Compliance can be described as the ability of a tubular structure to stretch and can be measured based on a percent change relationship between diameter and pressure. We design, build, and test a device that can measure the mechanical compliance of tubular structures. To demonstrate its utility, we have used vascular grafts as an example.

Millions of patients suffer from coronary and peripheral artery diseases. For treatment options, surgeons currently rely on autografts from one's own tissues, or synthetic grafts when autografting is not feasible.[2] Current synthetic grafts are not suitable for small-diameter blood vessel applications, where small-diameter blood vessels are defined to be less than six mm.[2] Among synthetic grafts, polyethylene terephthalate (PET), which can be described as Dacron©, or expanded polytetrafluoroethylene (ePTFE) are the most common choices in the clinic today. These grafts, however, are prone to aneurysm, infection, and

the potential for long-term performance and patency in applications requiring small diameters.

Part of the challenges in developing synthetic vascular grafts lies in identifying mechanically suitable material that can withstand cyclically changing pressures, stresses, and pulsations in physiological ranges while being flexible enough to experience changes in vascular compliance that do not result in permanent deformation or rupturing.[3] This unmet need has been addressed by designing synthetic vascular grafts that are altered physically, chemically, and mechanically to improve in vivo performance and mimic the mechanical and elastic behaviors of a native blood vessel.[4] Both solution electrowriting and electrospinning are methods to produce synthetic grafts by precisely patterning elastomeric fibers to optimize mechanical stability and structural integrity.[5]

Mechanical testing of graft material at physiological pressures is therefore crucial in characterizing its viability. Vascular compliance is especially important, as compliance mismatches with native blood vessels are related to processes that result in intimal hyperplasia.[6] Compliance values of native blood vessels also provide a means of comparison with synthetic vascular grafts, allowing researchers to understand whether a synthetic graft has mechanical properties similar to a native blood vessel.[4] For example, the compliance value of the femoral artery in humans can vary between 3.8% to 6.5%/100 mmHg while the compliance value of the saphenous vein in humans is typically 1.5%/100 mmHg.[7] Although previous methods for compliance testing have been developed, many of these methods are limited due to their lack of customization, reliance on external sources or companies, and use of expensive lab equipment.[8–11]

This paper describes a low-cost, user-friendly, and open-source data acquisition/analysis system that allows continuous recordings of pressure signals, diameter readings, and time points, which are used to calculate vascular compliance. This low cost assumes that one already has access to a laser micrometer. This measurement system is currently suitable for non-porous tubular structures and it simulates physiological ranges by inducing cyclic fluid pressure changes based on responses from a submersible water pump, pressure transducer, and a ball-valve.

METHODS AND PROCEDURES

Compliance Calculation Theory

This project draws on the supplementary methods of Wu, Wei to calculate vascular compliance.[12] As noted in the introduction, compliance can be calculated as a percent change in diameter between systolic and diastolic pressures. In trial 1, the systolic pressure was selected to be 150 mmHg and the diastolic pressure was 75 mmHg. In trial 2, the systolic pressure was 160 mmHg, and the diastolic pressure was 75 mmHg. These values were chosen to represent native blood pressures at high or systolic and low or diastolic pressures for arterial grafts. Compliance (%/100 mmHg) was calculated using **Equation 1** where D_{high} is the graft diameter at systolic pressures, D_{low} is the graft diameter at diastolic pressures, P_{high} is the systolic pressure, and P_{low} is the diastolic pressure. The diameter may be in any unit, but the pressure must be in mmHg. The 10,000 functions to convert to the % 100 mmHg units. In general, the calculation of the compliance of tubular structures will take the same form as **Equation 1** but will differ in the pressure and diameter ranges. For example, fluid suspension of food particles slowly flows through the intestine and nonetheless gives rise to compliance values based on a pressure-diameter relationship.

$$C = \frac{D_{high} - D_{low}}{D_{low}(P_{high} - P_{low})} * 10{,}000 \qquad \textbf{Equation 1}$$

Water Flow Diagram

As the density of water is nearly equivalent to the density of blood, this project used water flow to mimic the movement of blood through a native blood vessel.[13] The medium of flow can differ in other tubular structures and may require submersible pumps that can flow such mediums. A simplified diagram that only depicts the path of water flow through the compliance testing device is shown in **Figure 1**. Water first flows from the submersible water pump to the pressure sensor, the graft, and finally the ball valve. The ball valve is normally closed. The ball valve opens once it reaches the systolic pressure of 150 mmHg in trial 1 and 160 mmHg in trial 2. This opening of the ball valve allows for a decreased build up in pressure and thus graft compression. The ball valve closes at pressures below 75 mmHg. This closing of the ball valve allows for high pressure to build up within the graft so that it may experience expansion. The water eventually flows back into the reservoir once the ball valve is open. The process is repeated for cycles as water exits from the submersible water pump. This whole process results in diameter changes.

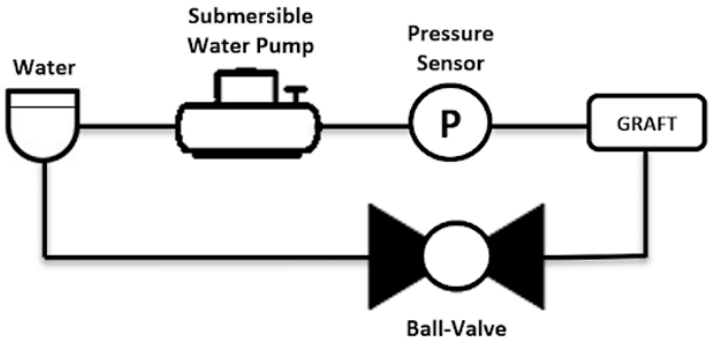

Figure 1. Piping and instrumentation diagram (PI&D) of water flow

Hardware Description

The measurement system consists mainly of electrical hardware as shown in **Figure 2**. The measurement system is powered through a 5 V power supply to the Arduino Uno and 12 V power supply to the ball valve. The system output is transmitted via a serial monitor using a USB, allowing for integration into other systems or software. In this project, the serial communication is read through PUTTY, which will be described in the *Software Description*.

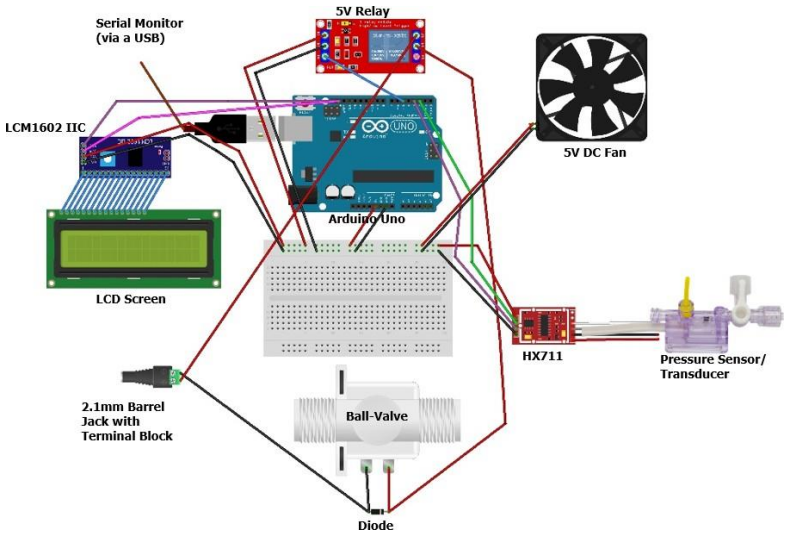

Figure 2. Wiring diagram of the pressure measurement component created with fritzing.

Each electrical component carries out a specific function: Arduino uno controls circuit components according to computer code; ball valve controls the flow of water and the build-up of pressure within the device; Edward Truwave pressure transducer relays the pressure information in mmHg and transmits data to Arduino; HX711 weighing sensor module increases the strength of the signal from the pressure transducer (e.g., the pressure transducer signal is in microvolts and Arduino only reads the signals in volts); 5 V relay module acts as a switch that turns off and on the power supply to the ball valve; LCD screen functions as an external display system for water pressure in mmHg; LCM1602 IIC acts as an adaptor that simplifies LCD wiring to the Arduino; barrel jack acts as a plug for 12 V power supply; 5 V DC fan functions as a ventilator to prevent potential overheating of the device; diode functions to prevent frying of relay circuitry, voltage spikes, and back emf.

The reduced bill of materials with key components is listed in **Table 1**. The measurement system offers a low-cost, customizable, and open-source option for compliance measurement. This system is low-cost, assuming one already has access to a laser micrometer.

Component	Quantity	Total Cost (USD)	Vendor	Part Number
Arduino Uno	1	28.5	Arduino	A000066
Motorized Ball Valve ¼"	1	34.9	U.S. Solid	JFMSV00027
Truwave Pressure Transducer	1	34.95	Edwards-Lifesciences	PX6001
HX711 Load Cell Amplifier	1	3.7	DIYmall	FBA_FZ0728X2
5V One Channel Relay Module Relay Switch	1	3.7	HiLetgo	3-01-0340
5V DC Cooling Fan	1	3.9	Aokin	8010
I2C 1602 LCD Display Module	1	5.99	GeeekPi	B07S7PJYM6
12V DC Power Connector (Barrel Jack)	1	2.99	Posdou	819356
480GPH Submersible Pump (1800L/H, 25W)	1	16.99	VIVOSUN	N/A
Solderless Prototype PCB Breadboard	1	0.99	Deeoee	7545924028
Total Cost Key Components		136.7		

Table 1. Reduced bill of materials with key components/totals only. It should be noted that the laser micrometer is not included in this calculation.

HX711 Load Cell Amplifier & Edwards-Truwave

The HX711's primary purpose is to amplify signals from pressure transducers or load cells and transmit these signals to a microcontroller such as the Arduino Uno. To obtain a pressure signal, the HX711 functions as a differential amplifier to get the difference between two voltage pins. This is necessary since the pressure transducer reads in the microvolt range (sensitivity 5.0μV/V/mmHg ± 1%) whereas the Arduino Uno reads in the volt range.

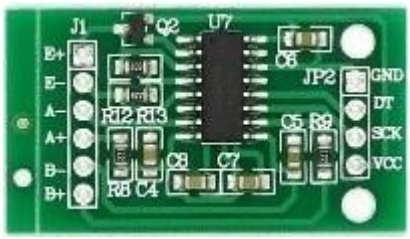

Figure 3. HX711 - Load Cell Amplifier 24-Bit ADC Converter. Image taken from Aerial Net.[14]

The HX711 load cell amplifier and its pins are shown in **Figure 3**. To acquire a signal, 4 of the 5 output pins are needed on the Edwards Truwave pressure transducer. These outputs are VCC, GND, SIGNAL+, and SIGNAL-, where pressure is proportional to (SIGNAL+ minus SIGNAL-), as shown in **Figure 4**.

At the start of the project, the pressure transducer appeared to be a strain gauge-type bridge with four connections. The fifth connection was likely a screen for a monitor cable. The output locations were proprietary information. However, the exact output locations were identified using resistance measurements between pairs of wires (e.g., V+ and V- or Signal+ and Signal-). Later, the company's patent confirmed the connections.[15]

Figure 4. Pin map for Edwards-Pressure Transducer.

The electrical connections between the HX711 and Edwards-Pressure Transducer are displayed in **Figure 5**. The pressure transducer pins V+ (5V), SIGNAL+, SIGNAL-, and V- (GND) are connected to on E+, A+, A-, and E- on the HX711.

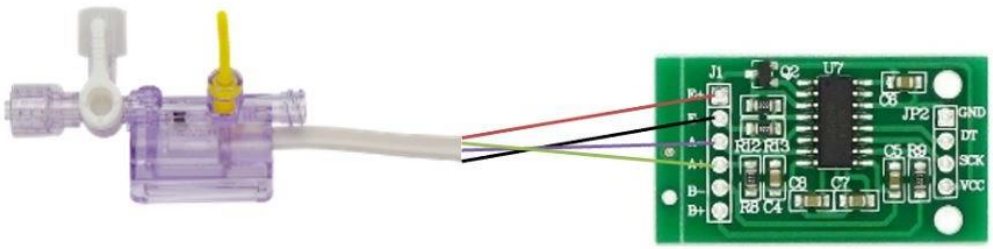

Figure 5. Pin map for Edwards-Pressure Transducer to HX711.

The HX711 pins GND, DT, SCK, and VCC have standard connections to the Arduino Uno. GND was connected to the ground on the breadboard, VCC was connected to the 5 V on the breadboard, SCK was connected to digital pin 2, and DT was connected to digital pin 3. Finally, the default sample rate of the HX711 is 10 Hz. To increase the sampling rate to 80 Hz, a hardware change on the HX711 must be made.[16] It is necessary to increase the sample rate of the HX711 to 80 Hz because a higher sample rate provides for more data points over a given period, allowing for a finer resolution of the data. In addition, a higher sampling rate allows for the capturing of subtle changes in data that may be missed at a lower sampling rate, especially when time-sensitive measurements are taken.

3D-Printing (Fusion Autodesk)

An enclosure was created with Autodesk Fusion to surround the electrical components and the Arduino Uno. Gates were made for the USB connector, 12 V DC Power Supply, pressure sensor wires, ball valve wires, and LCD. Additionally, a screwable lid was created to fix the enclosure. The cable grip glands were super-glued in the enclosure to prevent tugging of the pressure transducer. The 12 V DC barrel jack was superglued to prevent the movement of the power supply components. Lastly, both the LCD and the relay switch were screwed onto the enclosure. The print orientation for the 3D-printed enclosure and cover are shown in **Figure 6** and **Figure 7**. The final 3D-printed enclosure and lid configuration are shown in **Figure 8**.

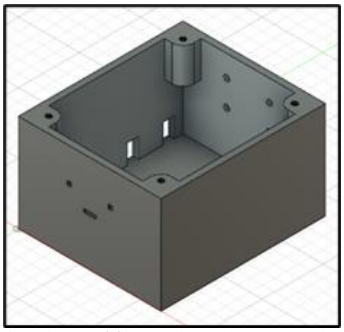

Figure 6. Print orientation for the 3D-printed enclosure. The enclosure is 117 mm in length, 135 mm in width, and 79.763 mm in height.

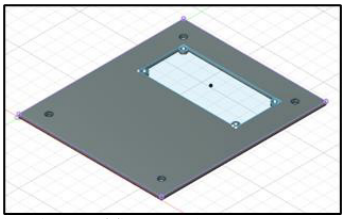

Figure 7. Print orientation for the 3D printed cover. The cover is 117 mm in length, 135 mm in width, and 2.381 mm in height.

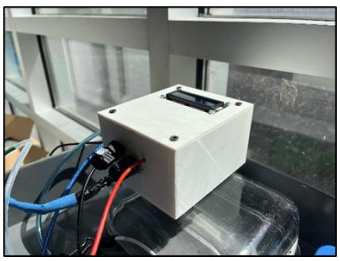

Figure 8. Final 3D-printed enclosure and lid configuration.

Hardware Build Instructions (Summarized)

1. Order an Edwards-Truwave Pressure Transducer and the wiring components listed.
2. Print the enclosure and cover lid using a 3D printer and the associated Autodesk Fusion files.
3. Isolate the top four wires on the Edwards-Truwave Pressure Transducer. Cut the plastic between each wire so that each wire pin is isolated from the other.
4. Perform a hardware change as indicated in the section of HX711 (Sampling Rate), which will increase the sampling rate to 80 Hz.
5. Solder the wires of each pin to the HX711 using the configuration in Figure 5 and then connect the HX711 to the Arduino Uno.
6. Calibrate the Edwards-Truwave Pressure Transducer using a baumanometer as indicated in the Pressure Calibration section.
7. Connect the hardware components as indicated in the Wiring Diagram section.
8. Glue the breadboard, Arduino Uno, and 12 V barrel jack to the bottom of the 3D-printed enclosure.
9. Screw the LCD to the cover lid and the 5 V relay module to the enclosure.
10. Glue the four wires of the Edwards-Truwave Pressure Transducer while leaving a gap between each wire.
11. Electrical tape the four wires to prevent potential tugging and damage of the wires.
12. Glue the grip cable glands into the 3D-printed enclosure. Connect the Edwards-Truwave Pressure Transducer wires to the grip cable gland and then tighten the grip cable gland.
13. Assemble the cover lid and enclosure by screwing both parts together.

Software Description

The Arduino Uno is an open-source microcontroller that processes analog and digital inputs and output pins. In this project, the Arduino Uno functioned to process signals from the Edwards-Truwave Pressure Transducer and display these reads through the serial monitor/plot (e.g., pressure versus time). It also controlled the LCD and ball valve as well as other electrical components listed in the Wiring Diagram subsection. The programming libraries used in this project include HX711.h, Wire.h, and LiquidCrystal_I2C.h. The program first calculated the water pressure, displayed the pressure on the LCD, opened/closed the ball valve based on water pressure, and then repeated the process. The LCD is mainly an accessory that is used in the scenario where one does not wish to look at the serial monitor. The LCD refreshes the pressure value displayed every time the void loop is executed in Arduino Uno. In other words, every time the water pressure is calculated. Therefore, the serial monitor is the best for viewing the data and its changes over time. PUTTY was used with Arduino to read data from the serial monitor and save it onto a .csv file. LabChart Pro was used to record diameter readings versus time and save them onto a .xlsx file. To account for differences in sampling rates and time delays, a MATLAB application processed and synchronized both signals based on the identification of peaks. The data acquisition/analysis workflow is shown in **Figure 9**.

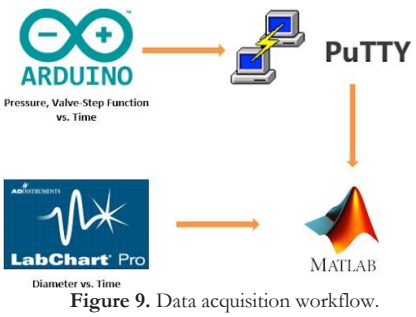

Figure 9. Data acquisition workflow.

The full operation/device protocol, programming code, and instructions are provided on GitHub and Google Drive.[17, 18]

Operation Instructions (Summarized)

1. Setup the system by connecting the submersible water pump, pressure sensor, vascular graft, ball valve, and laser micrometer as indicated in **Figure 1** and **Figure 10**.
2. Load the Arduino Uno code, serial_plott_water_valve_and_TruWave_pressure_sensor_HX711_lcd_.ino, onto the Arduino Uno. The Arduino USB cable should be connected to the Arduino Uno.
3. Load the LabChart Pro program file, 20180627_CETS_Compliance.adiset, onto LabChart.
4. Plug in the 12 V DC power supply to the barrel jack.
5. Place the submersible water pump inside the water reservoir and connect the plug to a power supply.
6. Record pressure signals and time points by PUTTY.
7. Record diameter readings and time points by LabChart.
8. Save the Arduino data files as a .csv and the LabChart data files as a .xlsx.
9. Load the MATLAB code, compliance_data_sync.m, onto MATLAB.
10. Rename the file variables in the MATLAB program according to the saved file names.
11. Run the MATLAB program and save the output files. A video of the measurement system and photographs can be viewed in Google Drive.[19]

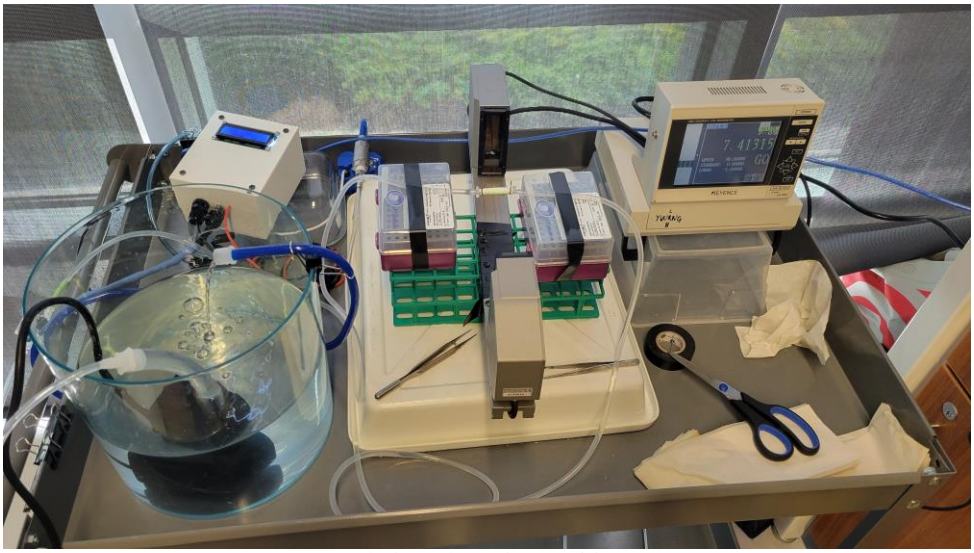

Figure 10. Final experimental setup of the measurement system.

RESULTS

Pressure Calibration (Edwards Truwave Pressure Sensor)

The pressure transducer was calibrated using a baumanometer that reads from 0 mmHg to 300 mmHg. The calibration curve was obtained from the pressure gauge and load signal (ADC value) readings. The ADC value is a digital signal read from Arduino, which corresponds to a pressure signal. The pressure signals were recorded in increments of 20 mmHg with the corresponding ADC value. The range for pressure acquisition is from 0 to 300 mmHg. The pressure versus ADC value plot is shown in **Figure 11**.

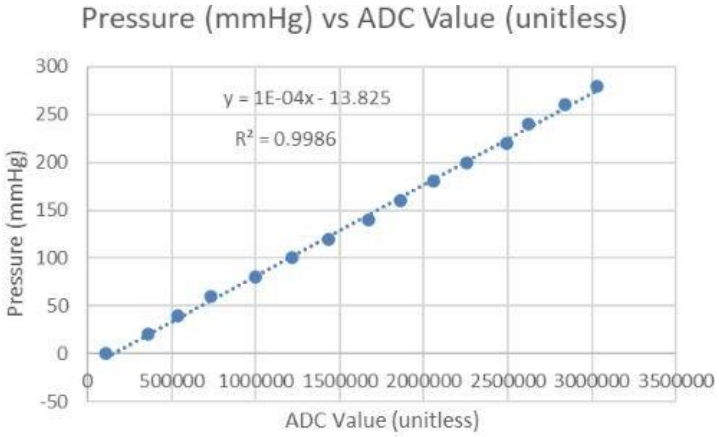

Figure 11. The ADC value recorded by the Arduino Uno and the corresponding pressure values in mmHg. The R2 coefficient shows a strong linear correlation and the regression line is used to calibrate the device.

The coefficient of determination of the linear least-squares fitted the data, R2 > 0.99, suggesting very strong linearity and minimal noise.

Pressure, Diameter, and Compliance versus Time Plots

A graph of pressure versus time, diameter versus time, and synchronized data with compliance values are shown in **Figure 12**, **Figure 13**, and **Figure 14** below. For all plots, the green stars represent the peak at either a maximum or minimum. A second trial run was performed using a graft of different material properties. The plots can be viewed in **Figure 15**, **Figure 16**, and **Figure 17**. All of the plots and the associated data can be viewed in Google Drive.[20]

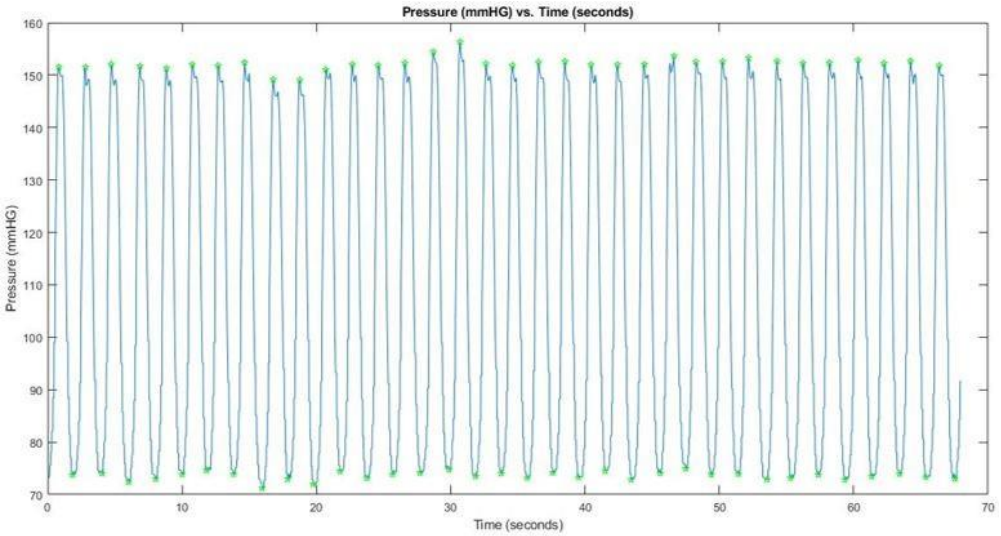

Figure 12. A plot of pressure signals in mmHg versus time points in seconds plots generated from MATLAB for the first trial run.

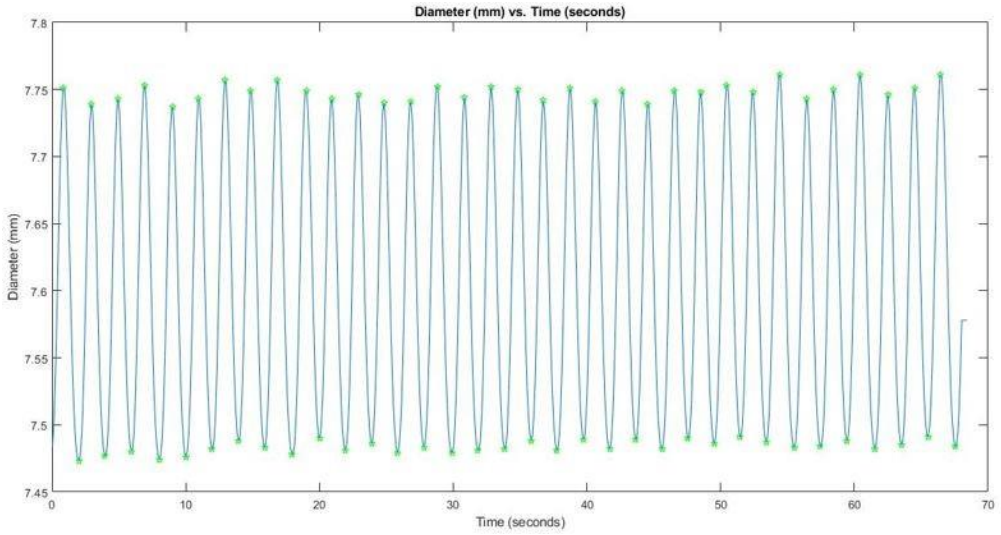

Figure 13. A plot of diameter readings in mm versus time points in seconds generated from MATLAB for the first trial run.

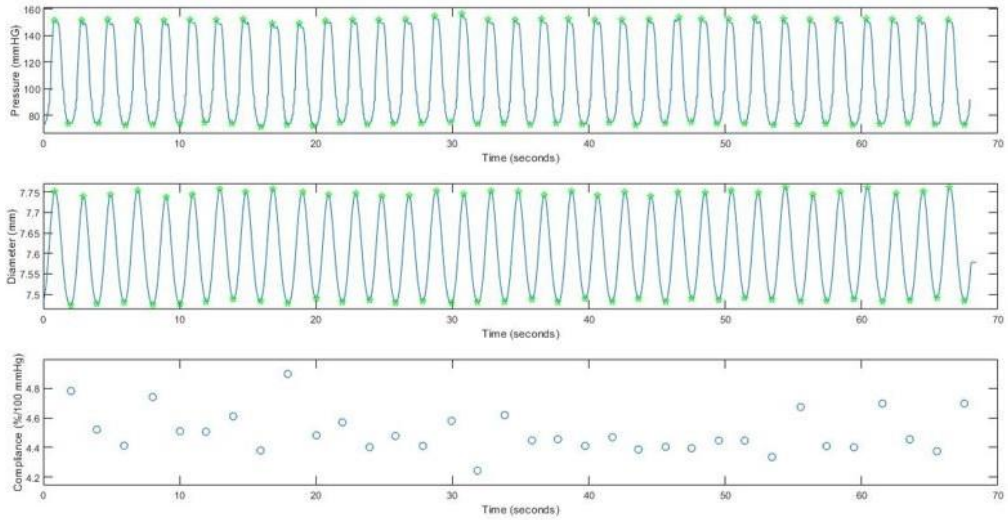

Figure 14. A plot of synchronized pressure, diameter, and compliance readings generated from MATLAB for the first trial run.

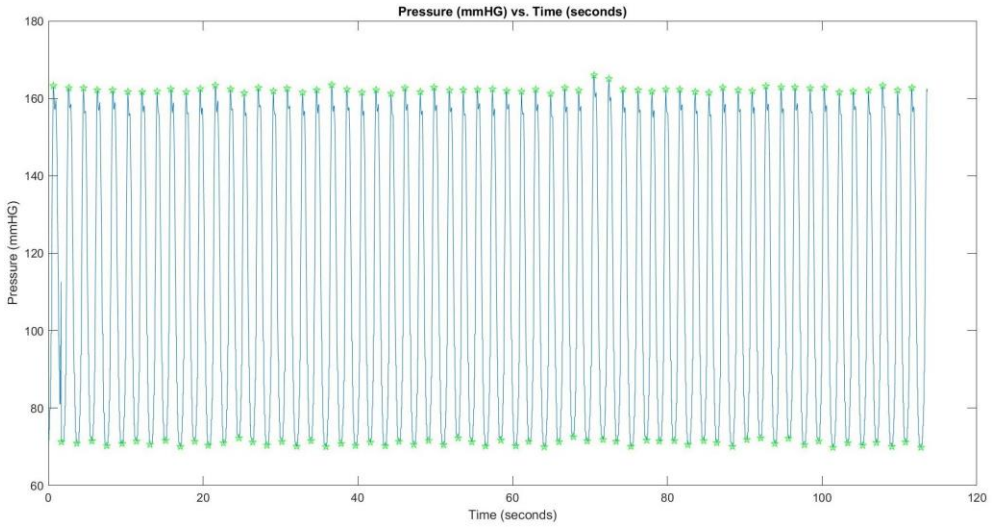

Figure 15. A plot of pressure signals in mmHg versus time points in seconds plots generated from MATLAB for the second trial run.

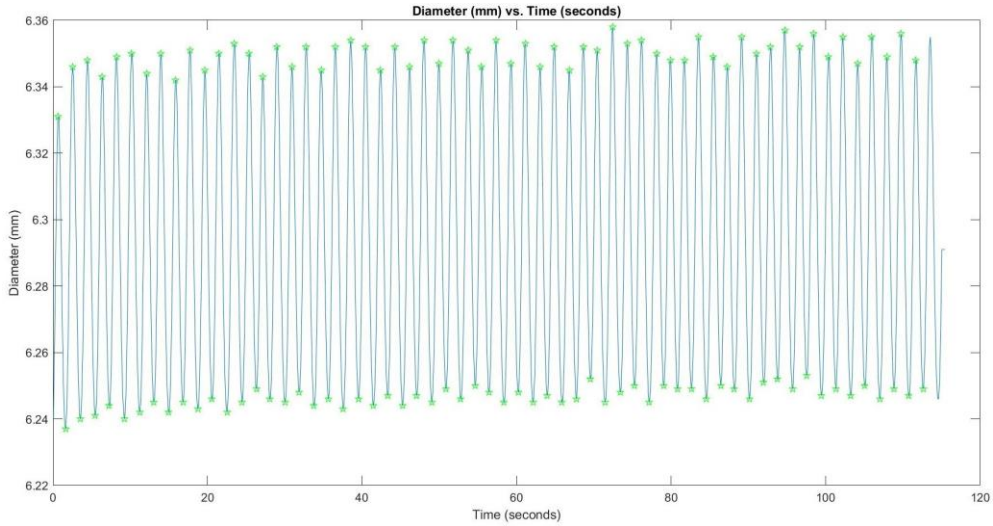

Figure 16. A plot of diameter readings in mm versus time points in seconds generated from MATLAB for the second trial run.

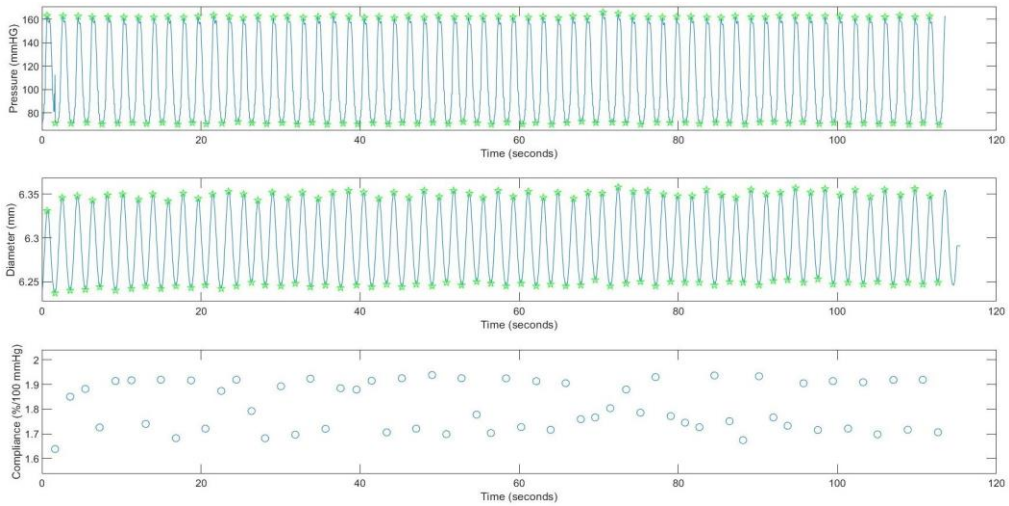

Figure 17. A plot of synchronized pressure, diameter, and compliance readings generated from MATLAB for the second trial run.

DISCUSSION

The main objective of the project was to develop a measurement system for compliance in non-porous tubular structures with vascular grafts as an example. The vascular grafts used in the project were arterial grafts though they were not optimized to match arterial compliance. The grafts functioned as a test sample for the measurement system. The system's significance lies in its ability to acquire compliance for non-porous tubular structures of varying material properties and dimensions. The system successfully measured pressure signals and diameter readings and subsequently calculated compliance. The Edwards Pressure Transducer used in this project was able to measure small changes in pressure signals and relay that information to the Arduino Uno, as it is a reliable transducer used in other projects.[21] This system was operational throughout the entirety of the trial runs. Each trial produced graphs of similar waveforms, particularly pressure signals and diameter readings that demonstrated oscillatory behavior as expected. Both trials 1 and 2 produced plots of similar behavior, indicating the device's consistency in measuring and computing data with no dependence on the tubular structure at hand.

The MATLAB program successfully synched different datasets by calculating time delays and identifying peaks as maximums or minimums in both the pressure versus time plot and diameter reading versus time plot as indicated in **Figure 12**, **Figure 13**, **Figure 14**, **Figure 15**, **Figure 16**, and **Figure 17**. The MATLAB program accounted for possible irregularities in data collection by introducing a five-second delay in recording data.

The system's operational range for pressure was sufficient to mimic physiological conditions. The main determinant of the pressure attained was the ball valve. The ball valve functioned to respond to pressure signals based on a program that opens or closes it once certain pressure thresholds are reached. It should be noted that a high enough flow rate on the submersible water pump is needed for pressure to be generated. As shown in **Figure 1**, the submersible water pump (VIVOSUN 480GPH Submersible Pump) releases a high stream of water that leaves the pump, hits the closed ball valve, and thus leads to a buildup of pressure. The pressure is relieved and subsequently decreased once the ball valve is opened at a certain threshold. Therefore, the ball valve allows for pressure to build up or to be relieved while the submersible water pump allows for a constant stream of high-pressure water to move within the system. The ball valve sets the pressure range based on the threshold programmed in Arduino Uno, which can only be achieved if there is a high enough flow rate on the submersible water pump.

The current VIVOSUN 480 GPH Submersible Pump had a maximum pressure of 170 mmHg. The VIVOSUN Submersible Pump has 210 GPH, 400 GPH, or 660 GPH versions that could lead to different pressure ranges depending on the users' specific need.

It should be noted that during the study, the location of the submersible pump and the water reservoir was moved from the bottom shelf of the cart to the top shelf of the cart as shown in **Figure 10**. When the submersible pump and the water reservoir were at the bottom shelf, the pressure range plateaued to about 110 mmHg and did not reach a high enough threshold to open the ball valve, in contrast to the reservoir at the top shelf with a pressure range between 75 mmHg to 150 mmHg. This suggested that gravity played an effect, and thus, it is recommended that all system components are leveled on the same plane, including the tubing within the system.

The compliance versus time plots were calculated as shown in **Figure 14** and **Figure 17**. The plots showed that compliance varies with time but stays within a certain range. Thus, the best approximation of graft compliance would be the average of all individual

Unlike previous systems for compliance testing, the system described in the paper does not rely on external sources or companies and offers customization for future users. For example, the DynatekLabs' DCT Dynamic Compliance Tester verifies % radial compliance under FDA, ISO, AAMI, and ASTM testing guidelines, but it does not offer customization and requires outsourcing from the company to test vascular compliance.[8]

The Universal Testing Machine (MTS Systems Corporation) uses a 50 lb. load cell and displacement velocity to correlate linear force and displacement to compliance, but it does not test compliance of the graft under physiological fluid pressure ranges as the system described in this paper.[10]

The Cardiovascular Regenerative Engineering Laboratory Marquette-MCW has a vascular graft test station to ISO 7198 standards that can measure dynamical radial compliance and burst strength under physiological pressures but does not make use of Arduino hardware or software to customize different components.[22]

The system in the paper is not capable of measuring dynamic radial compliance to ISO 7198 standards.[23] The system was not tested to ISO 7198 standards. The ISO 7198 standard has a 50-70 beats per minute requirement for the cycle rate. The cycle rate of the system in seconds can be approximated by dividing the number of cycles by the total time, providing cycles per second. Beats per minute can be calculated by multiplying the cycles per second by 60. The cycle rate of the system was approximately 30 beats per minute where **Figure 12** shows 35 cycles within approximately 70 seconds and Figure 15 shows 60 cycles within approximately 120 seconds. The system in the paper had a fixed and non-adjustable cycle rate that was not hardcoded in Arduino. Although the cycle rate of the system was not tested to ISO 7198 standards, one possible solution to adjust the cycle rate would be to change the flow rate of the VIVOSUN submersible pump, which is currently at 480 GPH, by using other VIVOSUN submersible water pumps.

The ISO 7198 standard has a requirement that a system can be cycled between 50-90 mmHg, 80-120 mmHg, and 110-150 mmHg. The system in the paper was not tested to ISO 7198 standard and was primarily tested for aortic pressure ranges, specifically between 75-150 mmHg in trial 1 and 75-160 mmHg in trial 2. This range was chosen based on our test samples which were arterial grafts. Although the system was not explicitly tested for ranges in the ISO 7198 standards, we believe that is capable of operating within the ranges stated by making changes to the system. First, the minimum and maximum pressure ranges, as determined by closing and opening of the ball valve, are hardcoded into the Arduino Uno as H_limit for high water pressure limit and L_limit for low water pressure limit. If one keeps the flow rate of the submersible pump used in this paper constant (480 GPH), the system should be able to cycle between 80-120 mmHg and 110-150 mmHg by changing the H_limit and L_limit in the Arduino Code, especially since the largest pressure range tested was between 75 mmHg to 160 mmHg. The 50-90 mmHg pressure range would require additional tests of the system to confirm its performance.

A limitation of this device is that it was largely dependent on the pressure transducer. The data collected depends on the accuracy of the linear regression curve with respect to the pressure transducer's readings, so damage to the transducer would produce inconsistent results. Moreover, if one wishes to change the pressure transducer, then the transducer must be recalibrated. A second limitation was the space requirement, as the experimental setup requires a large lab space to account for the device components as shown in **Figure 10**.

CONCLUSIONS

The measurement system described in this paper presented a cost-effective and customizable option to obtain compliance with tubular structures. The results demonstrated that the system and its associated programs operate as intended, allowing researchers to compare tubular structures of different lengths and material properties. There is potential for future improvements in this measurement system. The device currently supports only non-porous tubular structures. The electrical components could be replaced with cheaper options that would still allow the device to function properly.

The 3D printed box's STL files could include holes for the cable grips or the barrel jack, which would remove the need to drill holes. Lastly, a flow rate sensor could be inserted into the device so that it may quantify flow rate and produce flow rate versus time plots. This would further standardize the protocol by offering another means of analysis.

ACKNOWLEDGEMENTS

The author thanks the Department of Biomedical Engineering at Cornell University for their support and lab space.

REFERENCES

1. Gora A., Pliszka D., Mukherjee S., and Ramakrishna S. (2016) Tubular Tissues and Organs of Human Body– Challenges in Regenerative Medicine, *J. Nanosci. Nanotechnol.* 16, 13–39. *https://doi.org/10.1166/jnn.2016.11604*
2. Stowell, C. E. T., and Wang, Y. (2018) Quickening: Translational Design of Resorbable Synthetic Vascular Grafts, *Biomaterials* 173, 76–86. *https://doi.org/10.1016/j.biomaterials.2018.05.006*

vascular prostheses, *Phys. Procedia* 21, 234–239. *https://doi.org/10.1016/j.phpro.2011.10.035*
4. Stoiber M., Grasl C., Frieberger K., Moscato F., Bergmeister H., and Schima H. (2020) Impact of the testing protocol on the mechanical characterization of small diameter electrospun vascular grafts, *J. Mech. Behav. Biomed. Mater.* 104. *https://doi.org/10.1016/j.jmbbm.2020.103652*
5. D'Amato, A. R., Ding, X., and Wang, Y. (2021) Using Solution Electrowriting to Control the Properties of Tubu- lar Fibrous Conduits, *ACS Biomater. Sci. Eng.* 7, 400–407. *https://doi.org/10.1021/acsbiomaterials.0c01419*
6. Mugnai D., Tille J. C., Mrowczynski W., Valence S. D., Montet X., Moller M., and Walpoth B. H. (2013) Experimental noninferiority trial of synthetic small-caliber biodegradable versus stable vascular grafts, *J. Thorac. Cardio- vasc. Surg.* 146, 400–407. *https://doi.org/10.1016/j.jtcvs.2012.09.054*
7. Suzan O., Janset O., Hande S., and Ipek Y. (2023) The Effect of Polymer Type and Fiber Orientation on the Compliance Properties of Electrospun Vascular Grafts, *Fibres, and Textiles* 30, 67–71. *https://doi.org/10.15240/tul/008/2023-1-011*
8. Dynatek Labs, DCT Compliance Tester Mock Artery Compliance Tester, *https://dynateklabs.com/compliance-tester- stent-testing/* (accessed June 2023)
9. Tai N. R., Salacinski H. J., Edwards A., Hamilton G., and Seifalian A. M. (2002) Compliance properties of con duits used in vascular reconstruction, *BJS* 87, 1516–1524. *https://doi.org/10.1046/j.1365-2168.2000.01566.x*
10. Johnson J., Ohst D., Groehl T., Hetterscheidt S. and Jones M. (2015) Development of Novel, Bioresorbable, Small- Diameter Electrospun Vascular Grafts, *J Tissue Sci Eng* 6. *http://dx.doi.org/10.4172/2157-7552.1000151*
11. Furdella K. J., Higuchi S., Behrangzade A., Kim K., Wagner W. R., and Vande Geest J. P. (2021) In-vivo assessment of a tissue engineered vascular graft computationally optimized for target vessel compliance, *Acta Biomater.* 123, 298–311. *https://doi.org/10.1016/j.actbio.2020.12.058*
12. Wu W., Allen R. A., and Wang Y. (2012) Fast-degrading elastomer enables rapid remodeling of a cell-free synthetic graft into a neoartery, *Nat. Med.* 18, 1148–1153. *https://doi.org/10.1038/nm.2821*
13. Vitello D. J., Ripper R. M., Fettiplace M. R., Weinberg G.L., and Vitello J. M. (2015) Blood Density Is Nearly Equal to Water Density: A Validation Study of the Gravimetric Method of Measuring Intraoperative Blood Loss, *J Vet Med.* 2015. *https://doi.org/10.1155/2015/152730*
14. Aerial.Net Communications, HX711 - Load Cell Amplifier 24Bit, *https://www.aerial.net/shop/product/161_- 196/1673/hx711-load-cel. amplifier.html* (accessed June 2023)
15. Google Patents, Disposable blood pressure transducer and monitor interface, *https://patents.google.com/patent/US7604602B2/en* (accessed June 2023)
16. Youtube, Fast PASTA: converting an HX711 module to 80 SPS, *https://youtu.be/0cxS-a837bY* (accessed June 2023)
17. Kludze A., A Measurement System for Compliance in Tubular Structures, *https://github.com/Akludze/A- Measurement-System-for-Compliance-in-Non-Porous-Synthetic-Vascular-Grafts* (accessed June 2023)
18. Kludze A., Device Protocol - Tubular Structures Compliance Testing Device (Finalized), *https://drive.google.com/file/d/1Vp7v7METTrrN35440dMpPDkKP6ha26OA/view?usp=sharing* (accessed June 2023)
19. Kludze A., Compliance Test Device Pictures and Videos, *https://drive.google.com/drive/folders/16vMRIVjhI6aPlwsfjq1TgwseIj30oRS5?usp=sharing* (accessed June 2023)
20. Kludze A., Compliance Data Testing, *drive.google.com/drive/folders/1MZlq0JU5pA4hUrMV2mgXI- 6BCbTrZHbA?usp=sharing* (accessed June 2023)
21. R., Rolfe A., Cameron C., Shaw G. M., Chase J. G., Pretty C. G. (2022) Low cost circulatory pressure acquisition and fluid infusion rate measurement system for clinical research, *HardwareX* 11. *https://doi.org/10.1016/j.ohx.2022.e00318*
22. MCW & Marquette Cardiovascular Regenerative Engineering Laboratory, Development and Implementation of Mechanical Testing Methods For Tissue-Engineered Vascular Grafts, *https://mcw.marquette.edu/biomedical- engineering/cardiovascular-regenerative-engineering-lab/facilities.php* (accessed June 2023)
23. Camasao D. B., D. Mantovani (2021) The mechanical characterization of blood vessels and their substitutes in the continuous quest for physiological-relevant performances. A critical review *Mater. Today Bio* 10. *https://doi.org/10.1016/j.mtbio.2021.100106*

ABOUT THE STUDENT AUTHOR

Ave Kludze is a recent graduate from Cornell University. He graduated with a bachelor's degree in biomedical/medical engineering with a concentration in molecular, cellular, and systems engineering. He also worked as an academic excellence workshop facilitator and teaching assistant in calculus for engineers, multivariable calculus for engineers, differential equations for engineers, linear algebra for engineers, and biomedical transport phenomena. He plans to continue to conduct research in the field of cardiovascular biology and/or cardiology.

PRESS SUMMARY

Tubular structures are common in various lifeforms and organ systems. The device described here has applications for testing how a tubular structure responds to different pressures, using a vascular graft as an example. These synthetic grafts must be fabricated so that they are physically, chemically, and mechanically comparable to native blood vessels. To assess a graft's viability mechanically, vascular compliance testing is performed at varying physiological pressures. This is important since

autografts from one's own tissues, or synthetic vascular grafts when autografting is not feasible. This project focuses on a measurement system to characterize compliance in non-porous tubular structures based on an Arduino-control system, laser micrometer, and data analysis/acquisition programs.

Fibroblast Embedded 3D Collagen as a Potential Tool for Epithelial Wound Repair

*Claire Behning[a1], Lia Kelly[a1], Emma Smith[b1], Yizhe Ma[a], & Louis Roberts[a]**

[a]Department of Biology and Biotechnology, Worcester Polytechnic Institute, Worcester, MA
[b]Department of Biomedical Engineering, Worcester Polytechnic Institute, Worcester, MA

[1]Authors contributed equal intellectual effort to this manuscript; order determined alphabetically.

https://doi.org/10.33697/ajur.2024.107

Students: cebehning@wpi.edu, liakelly357@gmail.com, eesmith42@outlook.com
*Mentor: laroberts@wpi.edu**

ABSTRACT
Collagen is a functional biomaterial with many applications, including wound healing. 3D collagen hydrogels mimic an *in vivo* cell culture experience in cell survival and growth studies. In experimentally examining human cells under contact with 3D collagen, it is possible to understand the role of collagen in human epithelial tissue repair. This study explored the growth and attachment response of human MCF-7 cells when exposed to 3D collagen by investigating if the presence of NIH/3T3 fibroblasts embedded within the collagen should produce an increased wound-healing response. 3D collagen and fibroblast presence were able to be analyzed in tandem with a "sandwich-like" configuration of the gels to determine how these variables impact or improve the tissue repair response in MCF-7 cells. Examinations in growth, attachment, viability, and migration patterns demonstrated that MCF-7 repair response may be increased when in contact with NIH/3T3 embedded 3D collagen without impairing viability. Most notably, results from the migration assay revealed that MCF-7 cells migrate the most when covered by and adhered to cellular 3D collagen. Fibroblast-embedded collagen on top of and below MCF-7 cells exceeded quantitative assessment to near confluency, whereas less than 50 counted cells per image migrated without any top collagen layering. The continuation of these methods could involve in vivo experiments that incorporate live animal models to determine if these results will continue to extend to live tissue.

KEYWORDS
Collagen; 3D Collagen; Fibroblasts; Wound Healing; Hydrogels; Tissue Repair; Migration

INTRODUCTION

Fibroblasts are structural mesenchymal cells in connective tissue that are fundamental to producing collagen fibers, proteoglycans, fibronectin, glycosaminoglycans, and other extracellular matrix (ECM) components.[1] In the wound healing process, the body's response phases are as follows: hemostasis, inflammation, migration, proliferation, and remodeling.[2] Fibroblast cells, during the proliferative phase, increase production of collagen, elastin, proteoglycans, and hyaluronic acid, and these molecules are reorganized into a new ECM during the final phase.[2] The presence of fibroblasts has shown increased keratinocyte migration in a 3D wound model compared to fibroblast absence, indicating the importance of these cells in tissue repair.[3] The NIH/3T3 fibroblast tetraploid cell line (ATCC® CCL-1658™) isolated from Swiss albino (*Mus musculus*) murine embryos is a widely used model system for their functionality as a transfection host.

Similarly, MCF-7 cells (ATCC® HTB-22™) are used in cell modeling for their unique wound-healing behavior. MCF-7 cells are epithelial human breast cancer cells isolated in 1973 from the pleural effusion of a 69-year-old woman.[4] Cancer cells hijack parts of the wound healing process to ensure survival.[5] 3D models can be used to identify invasive or migratory patterns compared with 2D models more clearly.[6] When using a 3D collagen model, cell migration and degeneration of the extracellular matrix can be observed, more accurately modeling cancer cell metastasis.[6] Because deregulated dynamics with the extracellular matrix are a main characteristic associated with cancers, treatments can be better directed because of this enhanced visualization of changes.[7] This

response can be paired with 3D matrices, which shows that when MCF-7 cells are cultured in hydrogels, there is up-regulation of breast-specific markers compared to a standard 2D culture.[8]

Collagen's triple helix structure, one of the most abundant in the ECM, provides an ideal scaffold for use with the above cell lines.[9, 10] Conventional 2D cell culture offers insight into cell pathology, physiology, and function. However, a 3D collagen matrix can more closely resemble *in vivo* interactions between cells.[11] Collagen's low cost and flexibility make it one of the most commonly used proteins in matrix creation. Additionally, pore size, ligand density, and stiffness are easily customizable when using collagen as a scaffold in 3D culture.

Different cell lines and biomaterials can be modified to deliver drugs or cell therapy, coat medical devices, and protect against infection.[12] The use of collagen in tandem with NIH/3T3 cells in medical applications is becoming more common for their excellent wound healing properties. Collagen is commonly used as a hydrogel material because of its biocompatibility and ability to promote cell attachment, proliferation, and migration. Cells have been cultured on and within collagen to mimic biological microenvironments and examine their potential for wound treatment.[13] One study seeded fibroblasts into a collagen-rich hydrogel; the hydrogel formulation was thought to be a candidate for diabetic wound healing.[14] Cell viability, proliferation, and migration were studied using *in vitro* models, which showed that fibroblast survival and proliferation increased with a higher collagen concentration.[14] Another study used a hydrogel made of recombinant human collagen as a skin graft overlay.[15] The hydrogels were subjected to various tests to determine their biological properties using cell lines and live rat subjects. A cytotoxicity study of HaCaTs, HUVECs, and primary human fibroblasts indicated that the collagen extract had slight cytotoxicity, particularly for the HUVECs.[15] A migration study indicated that the ability of each cell line to migrate was increased following six to 24 hours of culture.[15] Live animal experiments suggested that the hydrogel promoted wound healing, dermis formation, basement membrane formation, angiogenesis, and proliferation.[15] The researchers determined that this hydrogel has the potential to aid in wound healing as a graft overlay.[15] Collagen hydrogels with and without cells seeded within them have the potential to aid in the treatment of wounds, but further research and modeling are needed.

Much of the work described previously utilized 2D culture models and animal models to test a material's wound healing properties. However, there is more to be done in developing 3D culture models. Because of the biochemical healing properties naturally expressed by fibroblasts, we proposed that a hydrogel constructed from 3D collagen embedded with NIH/3T3 cells can potentially increase the wound repair response in MCF-7 epithelial tissue. Repair response was measured using physical and quantitative values of cell attachment, growth, viability, and migration. Observing consistent cell viability and increased attachment and migration, cellular collagen hydrogels have the potential to be used as a tool for tissue repair.

METHODS AND PROCEDURES

Cell Maintenance

NIH/3T3 (ATCC® CCL-1658™) and MCF-7 (ATCC® HTB-22™) cell lines from the American Type Culture Collection (ATCC) were maintained in complete medium containing Dulbecco's Modified Eagle Medium (DMEM) with 10% Fetal Bovine Serum (FBS) and 1% Penicillin/Streptomycin (PS). Cells were incubated at 37 °C and 5% CO_2 in T25 and T75 flasks. Cells were passaged every 48-72 hours as needed or when nearing confluence.

Cell Counting

Cells were trypsinized and diluted in complete medium. 100 µL of 0.4% Trypan Blue was added to a 100 µL aliquot of the cell suspension. 10 µL of the cell-dye mixture was loaded on a hemocytometer and analyzed using a 10x objective microscope. Cells within 1 mm^2 squares were manually counted, and the average of four squares was calculated. The total cell density was calculated using the following formula: average number of cells per square × 2 × 10^4.

Cell Suspension in 3D Collagen

Nearly confluent NIH/3T3 cells were trypsinized and resuspended in complete medium. For the attachment and proliferation and alamarBlue™ assays, a 3D collagen (Sigma-Aldrich® C4243™) mixture was prepared using 300 µL of cell suspension and 1.2 mL of collagen to create a final concentration of 2.02 × 10^3 cells per mL. For the migration experiment, the same method was used with different volumes: for the 300 µL and 200 µL aliquots, we used 360 µL and 240 µL of cell suspension with 1.44 mL and 960 µL of collagen mixture, respectively. The final concentrations were 3.9 × 10^3 cells per mL for the 300 µL aliquots and 4.6 × 10^3 cells per mL for the 200 µL aliquots. For the acellular collagen mixture, complete medium was added in place of a cell suspension. The collagen mixture was resuspended evenly. 300 µL and 200 µL aliquots were placed into the allotted collagen wells

for 12- and 24-well plates, respectively. The collagen was left to firm in a 37 °C incubator for two-to-three hours. Following gelation, MCF-7 cells were trypsinized to place on the set collagen gel. Medium was added to wells without cells seeded on top.

Attachment and Proliferation Assay

Attachment and proliferation were assessed by plating a 12-well plate of MCF-7 cells on acellular and cellular 3D collagen (**Table 1**). MCF-7 cells were seeded at equal concentrations on top of collagen either embedded with NIH/3T3 cells (Cell Col) or not embedded (Acell Col). Negative controls consisted of wells with collagen or on its surface. Positive controls consisted of wells with MCF-7 cells adhered and no collagen (No Col). 1 mL of complete medium was added to each well to ensure collagen hydration and cell maintenance. Observations were made after 24, 72, and 120 hours. At 120 hours, cell counts were performed and the remaining cells from representative conditions were transferred to a 96 well plate to perform an alamarBlue™ assay.

	1	2	3	4
A	MCF-7 Cells Only	MCF-7 Cells Only	Acellular Collagen with MCF-7s Seeded on Top	Cellular Collagen with MCF-7s Seeded on Top
B	Acellular Collagen	Acellular Collagen	Acellular Collagen with MCF-7s Seeded on Top	Cellular Collagen with MCF-7s Seeded on Top
C	Cellular Collagen	Cellular Collagen	Acellular Collagen with MCF-7s Seeded on Top	Cellular Collagen with MCF-7s Seeded on Top

Table 1. Design conditions for the attachment and proliferation assay in a 12-well plate.

AlamarBlue™ Assay

To assess viability of the MCF-7 cells in response to collagen exposure, an alamarBlue™ assay (BIOSOURCE® Cat. No. DAL1025) was performed using samples from a cellular control well, an acellular collagen well, and a cellular collagen well from the attachment assay. Each well was trypsinized and a cell count was performed with hemocytometry. The cell counts were found and used as a reference when setting up the assay. A 96-well plate was set up using three different samples: control MCF-7 cells grown in only medium (control), MCF-7 cells grown on acellular collagen, and MCF-7 cells grown on cellular collagen. A two-fold dilution series ranging from undiluted to 1:64 was performed, with a final cell solution volume in each well of 125 uL. 125 µL of medium was added to wells to serve as a blank. After incubating the plate for 72 hours, 12.5 µL alamarBlue™ Cell Viability Reagent was added to each well and incubated for two hours. A Biotek Synergy H1 Hybrid Multi-Mode Reader was used to read the fluorescence of the wells using excitation and emission values of 540 nm and 590 nm, respectively. The average of blank wells was calculated and used to find the blank-subtracted values of each well. Averages of the conditions (control, acellular collagen, cellular collagen) for each well dilution were calculated and used in statistical analyses using one-way ANOVA and two-sample t-tests ($p<0.05$).

Microscopy/Imaging

Each well was observed at 0, 24, 72, and 120 hours. Cells were assessed for attachment and growth through visual analysis. The shape of the cells, confluency, and other visually apparent changes were noted at each observation point. Images were taken through an Olympus IMT-2 Inverted Research Microscope at 100x magnification to record the qualitative data.

Migration Assay

Two 12-well plates of MCF-7 cells on acellular and cellular 3D collagen (**Table 2**). Alongside the 12-well plates, a 24-well plate was used for 200 µL collagen gels to be transferred on top of the scratch once firmed. MCF-7 cells were seeded on top of collagen either embedded with NIH/3T3 cells or not embedded. Two wells with MCF-7 cells adhered to the well bottom without collagen on top were scratched for positive controls. 1 mL of complete medium was added to each well to ensure collagen hydration and cell maintenance. Each well was then lightly scratched with a 1000 µL pipette tip to simulate a wound. An hour after scratching, the collagen gels were transferred to the 12-well plates and placed on top of the scratch. Plates were incubated at 37 °C and 5% CO_2. Following 24 hours, observations were conducted of each well, and images were taken using a Biotek Cytation 5 Cell Imaging Multimode Reader at 10x brightfield. Image recoloring was performed in Google. Scratch images were divided into three fields of view (FOVs) acting as three technical replicates per condition, and the number of cells migrated within the scratch were counted per FOV. The average and standard deviation were calculated for each condition, and both one-way ANOVA and two-sample t-tests ($p<0.05$) were performed to determine significance between groups.

	1	2	3	4
A	MCF-7 Cells Only with Acellular Collagen on Top	MCF-7 Cells Only with Acellular Collagen on Top	MCF-7 Cells Only with Cellular Collagen on Top	MCF-7 Cells Only with Cellular Collagen on Top
B	Acellular Collagen with MCF-7s Seeded on Top (Acellular Collagen on Top)	Acellular Collagen with MCF-7s Seeded on Top (Acellular Collagen on Top)	Acellular Collagen with MCF-7s Seeded on Top (Cellular Collagen on Top)	Acellular Collagen with MCF-7s Seeded on Top (Cellular Collagen on Top)
C	Cellular Collagen with MCF-7s Seeded on Top (Acellular Collagen on Top)	Cellular Collagen with MCF-7s Seeded on Top (Acellular Collagen on Top)	Cellular Collagen with MCF-7s Seeded on Top (Cellular Collagen on Top)	Cellular Collagen with MCF-7s Seeded on Top (Cellular Collagen on Top)

	1	2	3	4
A	MCF-7 Cells Only With No Collagen on Top	MCF-7 Cells Only with No Collagen on Top	EMPTY	EMPTY
B	Acellular Collagen with MCF-7s Seeded on Top (No Collagen on Top)	Acellular Collagen with MCF-7s Seeded on Top (No Collagen on Top)	EMPTY	EMPTY
C	Cellular Collagen with MCF-7s Seeded on Top (No Collagen on Top)	Cellular Collagen with MCF-7s Seeded on Top (No Collagen on Top)	EMPTY	EMPTY

Table 2. Design conditions for the migration assay in two 12-well plates. Wells contained a bottom layer of collagen, top layer of collagen, both layers, or neither. MCF-7 cells were seeded in all experimental wells.

RESULTS

Cellular 3D Collagen increases MCF-7 attachment and does not decrease cell viability.

Visual confirmation of gel conformation and cell attachment was completed prior to and immediately following cell seeding. In some cases collagen did not set evenly through the entirety of the well, or had air pockets (independent of the presence of cells). Every effort was made to acquire images in properly set and relatively unform collagen gels. **Figure 1** depicts the visual confirmation of cells being present following seeding. Immediately following seeding, cells maintained a rounded shape, indicating potential lack of attachment and spreading on the bottom of the well. This assay ran for 120 hours total. At the 24-hour time point of the assay, the MCF-7 cells seeded on top of the collagen had become partially attached. In all wells, a significant number of rounded cells remained, indicating a lack of attachment and spreading on the collagen gel. When examining for visual confirmation of attachment, cells appeared to be viable with typical conformation and growth patterns. The attempted monolayer creation of MCF-7 cells on the top of the collagen gel was successful, as all of the cells were visually in one layer and had not migrated or fallen into the gel layer. When observing the wells with NIH/3T3 cells embedded in the collagen, they were not easily visible, leaving only the MCF-7 cells to be viewed with a high level of certainty. A significant number of cells in every well lacked spreading and remained unattached or partially attached. Although medium was changed through pipette aspiration followed by addition of new media, many of the rounded cells remained. This suggests partial attachment rather than complete unattachment as aspiration would have removed fully unattached cells.

At the 72-hour time point (**Figure 2B**), each well was observed for cell conformation, attachment, viability, and growth. Based on visual observations, cells in the cell-only (No Col) well and on top of the acellular collagen (Acell Col) and cellular collagen (Cell Col), were all growing at an expected rate, reaching 40% confluency (**Figure 2B**). Some cells in every well continued to maintain a rounded shape indicative of lack of attachment or partial attachment and spreading. At the end of observations, each well was given fresh media.

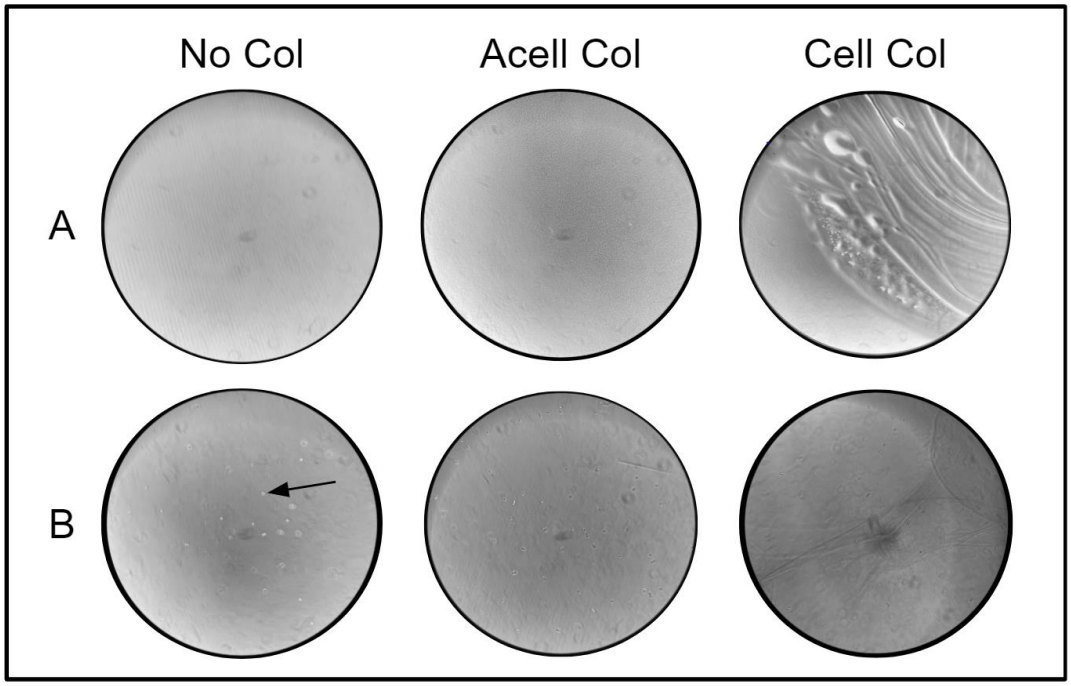

Figure 1. Attachment and Proliferation Assay, before **(A)** and after **(B)** cell seeding. From left to right: No Collagen (No Col), Acellular Collagen (Acell Col), and Cellular Collagen (Cell Col). Arrow indicates a rounded, not-yet or partially attached cell.

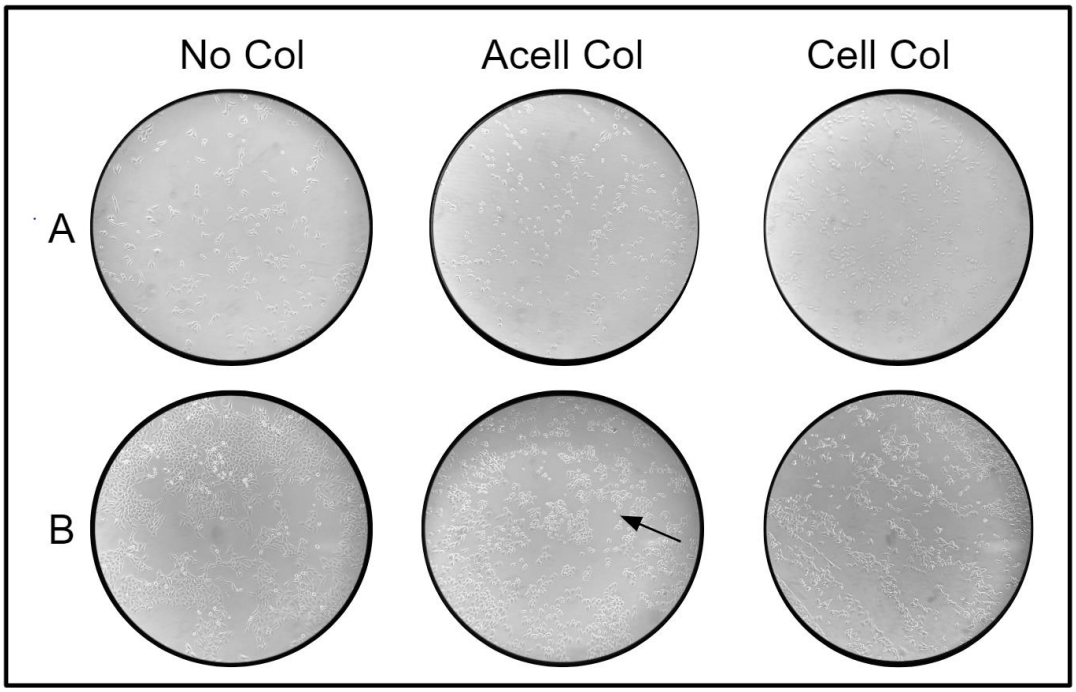

Figure 2. Attachment and Proliferation Assay, 24 **(A)** and 72 **(B)** hour time point. From left to right: No Collagen (No Col), Acellular Collagen (Acell Col), and Cellular Collagen (Cell Col). Arrow indicates a rounded, not-yet or partially attached cell.

Cells were again observed at the 120-hour time point for cell conformation, attachment, viability, and growth. Cells in all wells appeared to be more attached and spread than at the 72-hour time point. In wells with collagen that had MCF-7 cells seeded on top, distinct clusters of cells had been created. These areas had visual separations between highly confluent areas of cells and those with less confluence. This was noted on the wells with unseeded collagen and MCF-7 cells on top of the gel (**Figure 3A**).

The wells that appeared to have the highest confluency of MCF-7 cells were of the Cell Col condition. All cells appeared to be more attached than at previous time points and had less rounded conformation. Throughout each well, conformation and shape of cells was consistent with each other and with what is expected of the cell line. The average number of cells was 3.4×10^4 cells/mL in No Col control wells, 3.2×10^4 cells/mL in Cell Col wells, and 1.0×10^4 cells/mL in Acell Col wells. Cell counts revealed that there was no significant increase nor decrease in growth from the control condition compared to cellular collagen attachment. Interestingly, the decrease in cell concentration for acellular collagen conditions was remarkable. It appeared that cellular attachment and, most notably, confluency increased in collagen wells. In contrast, MCF-7 proliferation rates did not seem to improve as supported by the cell counts.

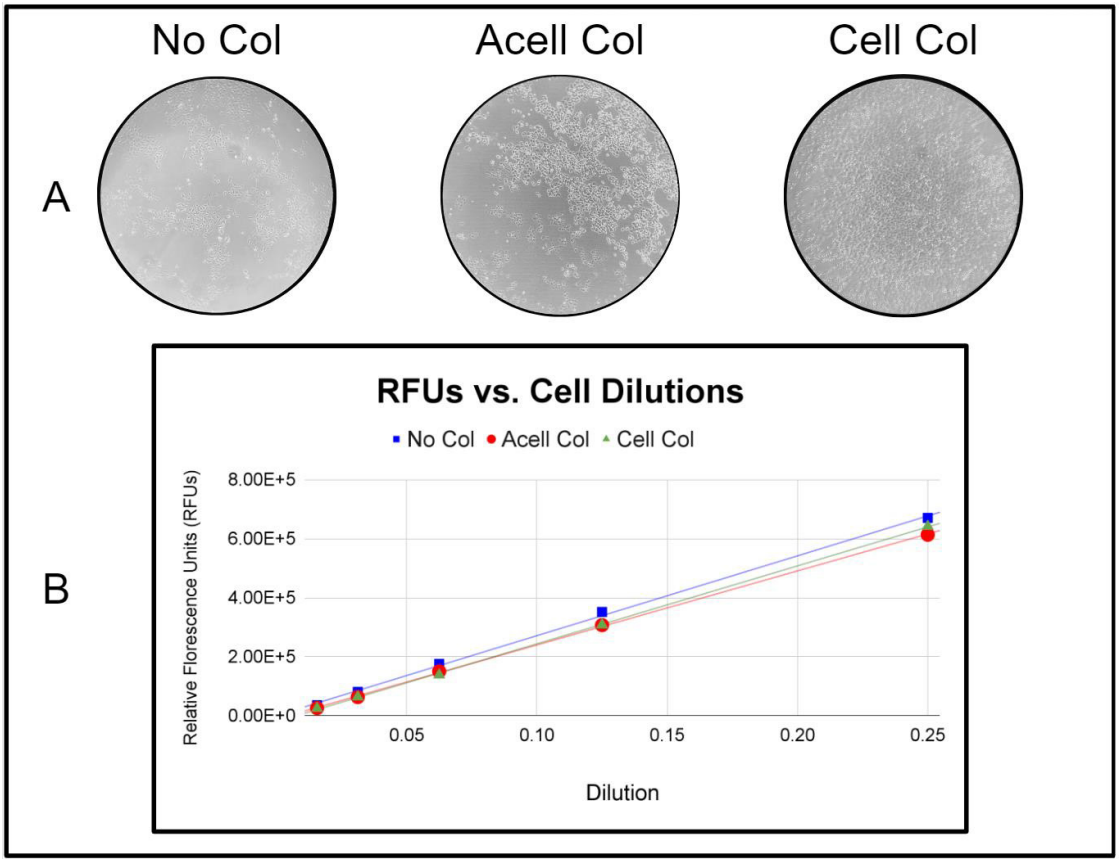

Figure 3. (A) Attachment and Proliferation Assay, 120 hour time point. From left to right: No Collagen (No Col, blue square), Acellular Collagen (Acell Col, red circle), and Cellular Collagen (Cell Col, green triangle). **(B)** Graph of fluorescence plotted against dilution for each condition of the previous attachment and proliferation assay.

The alamarBlue™ assay using a resazurin solution measured MCF-7 viability after no collagen exposure, exposure to 3D collagen, and exposure to fibroblast-embedded 3D collagen. When relative fluorescence units (RFUs) were graphed over dilutions, all three conditions showed a consistent linear growth across the dilutions 0.02 to 0.25 with R^2 values approaching 1 (**Figure 3B**). More concentrated dilutions (1.0 and 0.5) decreased the linear trend for each category, indicating that the concentration of cells in these wells exceeded detectable fluorescence levels. Therefore, data points from dilutions up to 0.25 were analyzed. Statistical analysis using both one-way ANOVA and two-sample t-tests comparing conditions in the 0.25 diluted wells revealed no significant differences ($p<0.05$), suggesting that contact with collagen does not affect cell viability.

MCF-7 cells migrate the most when covered by and adhered to cellular 3D collagen.

The migration assay began with plating MCF-7 cells onto either acellular or cellular collagen and incubating for 24 hours. At this time, cells not adhered to collagen were 30% confluent, cells adhered to acellular collagen were 75% confluent, and cells adhered to cellular collagen were 95% confluent (**Figure 4A**). After 24 hours of incubation, each well was scratched with a 1000 µL pipette tip (**Figure 4B**). Each scratch was made in the center of a well and did not puncture through the collagen gel, if present. After the scratches were made, collagen gels were added on top of designated scratches to represent a wound repair aid.

After a 24-hour incubation period, representative images of each scratched artificial wound were taken to assess the proliferation and migration of MCF-7 cells in response to 3D collagen treatment (**Figure 5**). The averages for each condition were calculated using three fields of view along the scratch (Figure 6A). Cell counts varied between conditions depending on the presence of collagen on top or bottom and whether fibroblasts had been embedded. Wells with acellular collagen on the bottom produced the highest attainable MCF-7 count (**Figure 6B**) when acellular collagen was placed on top, with an average of 78 migrated cells. Other wells contained more than 78 migrated cells but could not be counted accurately due to an overabundance of cells. All fields of view with cellular collagen on the bottom and top were too confluent to count the number of migrated cells.

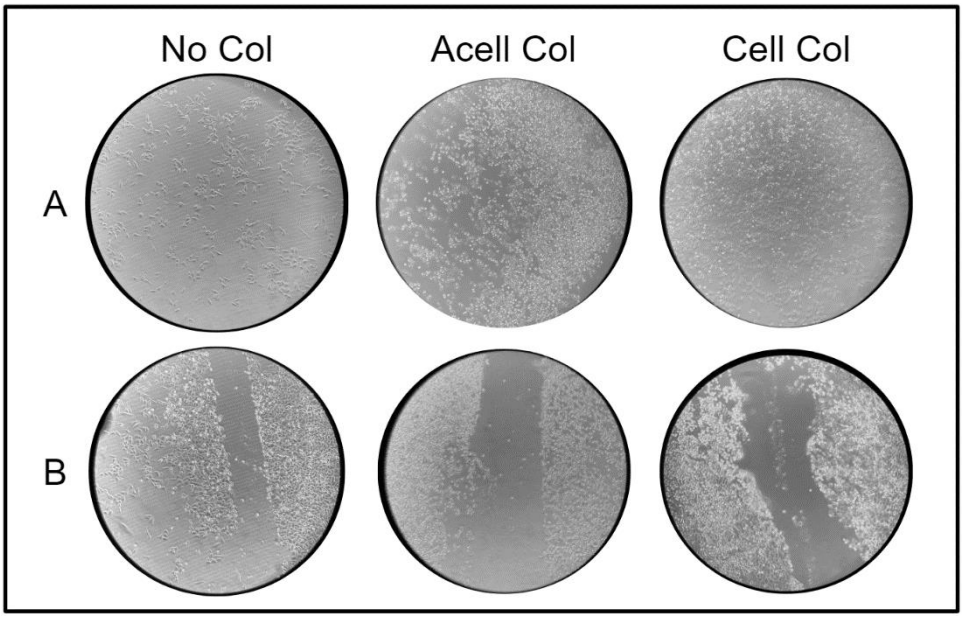

Figure 4. (A) NIH/3T3 Collagen base #2 MCF-7 cells before scratching. From left to right: Cells on No Collagen (No Col), MCF-7 cells on Acellular Collagen (Acell Col), and MCF-7 cells on Cellular Collagen (Cell Col). **(B)** Initial scratches of MCF-7 cells. From left to right: No Collagen (No Col), Acellular Collagen (Acell Col), and Cellular Collagen (Cell Col).

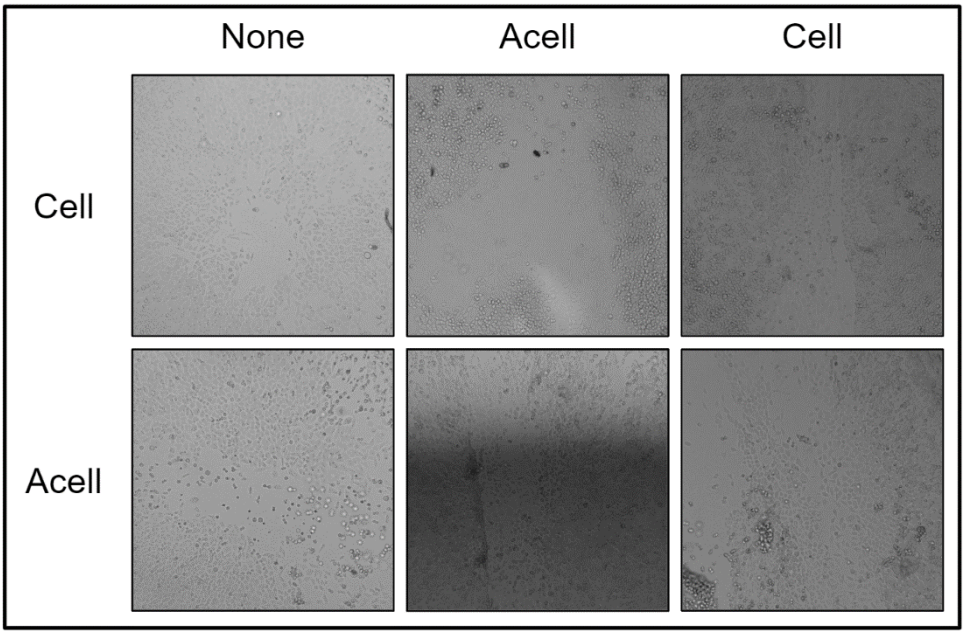

Figure 5. Scratches of MCF-7 cells following 24 hours (None - No Collagen on Bottom, Acell - Acellular Collagen on Bottom/Top, Cell - Cellular Collagen on Bottom/Top)

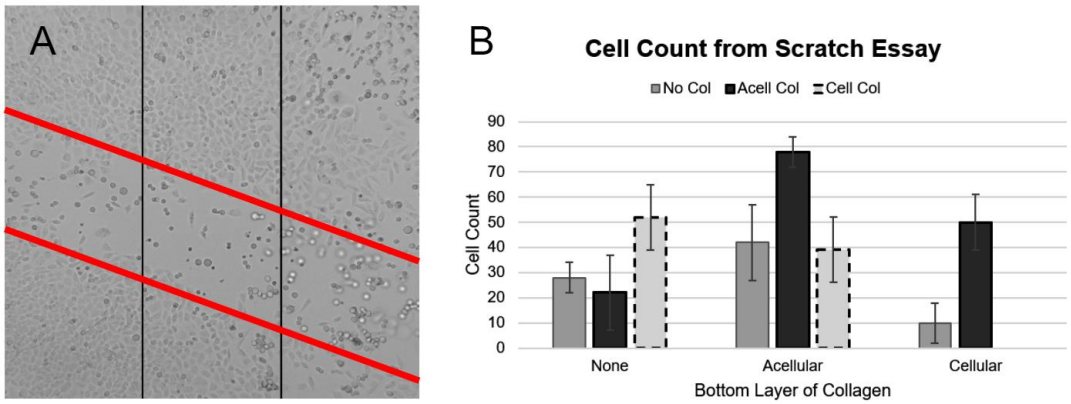

Figure 6. (A) Fields of View (FOV) for Cell Count, the black lines indicate the field of view and between red lines indicate the scratch location. Image taken at 100x magnification. **(B)** Results of Scratch Assay Cell Count Averages. Legend represents the condition of collagen on top. Error bars represent standard deviations of the technical replicates. The cell count for Cellular Collagen on Top/Bottom was not attainable due to an overabundance of cells to count.

The cells were microscopically examined, and images were taken after 24 hours (**Figure 5**) to perform a cell count of the scratched region. The wells with no collagen on the bottom had average cell counts of 28 ± 6, 22 ± 15, and 52 ±13 cells for the following conditions: no collagen on top, acellular collagen on top, and cellular collagen on top (**Figure 6B**). Of the conditions with no collagen on the bottom, the cellular collagen on top resulted in the largest number of migratory cells. The wells that had acellular collagen on the bottom had average cell counts of 42 ± 15, 78 ± 6, and 39 ± 13 cells for the following conditions: no collagen on top, acellular collagen on top, and cellular collagen on top, respectively (**Figure 6B**). The largest cell count was found in the well with acellular collagen on top and bottom. Finally, in the wells with cellular collagen on the bottom, the average cell counts were 10 ± 8 and 50 ± 11 cells for the conditions with no collagen on top and acellular collagen on top, respectively (**Figure 6B**). The condition where cellular collagen was present on the bottom and top of a cell monolayer was unable to be counted due to the density within the scratch.

Overall, conditions in which no collagen was placed on top of the scratch wound were associated with the least amount of cell migration. The conditions with the highest visible migration and wound healing rate were wells that had collagen placed on both the top and bottom with matched fibroblast variables (e.g., cellular collagen on top with cellular collagen on bottom). The wells with cellular collagen on both top and bottom or wells with acellular collagen on both top and bottom had the most growth success when compared to wells with no collagen placed on top of the wound or collagen that had mismatched cellular states. It is important to note that mere presence of collagen tended to increase migration rate, as the average counts for conditions without collagen on top were below 50 cells per FOV.

We performed one-way ANOVA statistical analyses to assess the significance of our results when greater than two conditions were being directly compared. The results of ANOVA did not meet the standard of significance; having data from more experimental replicates would likely increase the statistical resolving power. After performing a two-sample t-test ($p<0.05$) on the cell counts from the scratch assay in **Figure 6B**, the wells were compared to determine how significant cell migration was across scratch wounds. The migration of MCF-7 cells was larger and statistically significant for wells with acellular collagen on top and collagen on bottom, regardless of the presence of fibroblast cells within that bottom layer, when compared to wells that did not have any collagen layer on bottom (control). Similarly, wells with acellular collagen on top and collagen on bottom, regardless of the cellular status of that bottom layer, had a larger and statistically significant migration compared to wells with cellular collagen on top. Finally, cell migration was larger and statistically significant for wells with a top layer of cellular collagen, regardless of the presence of cells within the bottom layer, when compared to wells entirely without collagen and wells with only a top layer of acellular collagen.

DISCUSSION

This study aimed to explore the growth and attachment response of human MCF-7 cells when exposed to 3D collagen by investigating if the presence of NIH/3T3 fibroblasts embedded within the collagen should produce an increased response in cell repair. 3D collagen and fibroblast presence were analyzed in tandem with a "sandwich-like" configuration of the gels to determine how these variables impact or improve the tissue repair response in MCF-7 cells. Through the use of physical and quantitative values of cell attachment, growth, viability, and migration assay techniques, the wound repair response was measured and assessed. Examinations in growth, attachment, viability, and migration patterns demonstrated that MCF-7 repair response may be increased when in contact with NIH/3T3 embedded 3D collagen without impairing viability.

Based on the experimental plate setup with the collagen unevenly distributed, there could have been growth patterns of cells that were distributed or organized in a different orientation. This could have caused some of the stark lines or grouping of cell growth that we observed or the slower rate of attachment. Additionally, this environment and transfer likely caused a significant amount of cell stress which could have influenced the slowed attachment rate, more rounded epithelial shape, and decrease in proliferation for Acell Col wells. Overall, based on the repair pattern of the MCF-7 cells over time, the wells with the highest attachment rates were those with collagen gels seeded with NIH/3T3 cells. Despite this, conditions with cellular collagen were expected to have increased MCF-7 growth and replication due to NIH/3T3 cell wound healing capabilities. Fibroblast cells increase the production of collagen, elastin, proteoglycans, and hyaluronic acid, which promote new extracellular matrices to be developed. Further investigation should be done to confirm the absence of improved growth response from the cellular collagen-attached cells.

It is important that wound treatments do not impact cell viability nor cause increased cell death. Because collagen conditions did not impact viability compared to the control, we can infer that a collagen hydrogel does not produce any cytotoxic effects on live tissue cells. A majority of the dilution wells had linear trends with R^2 high values for the three conditions, indicating consistency in the viabilities. An ideal healing model would not only be safe and non-toxic to cells but also increase tissue regeneration. While these results did not show improved proliferation using a constant cellular concentration, further exploration of varying NIH/3T3 density within collagen may address this aspect.

When setting up the attachment and viability assay, multiple wells had collagen bases that did not set evenly through the entirety of the well. This was a factor that had to be considered through the whole assay as the inconsistency of the gel width could cause data discrepancies. Additionally, the collagen appeared to have pockets of air when observed with a microscope before seeding cells (**Figure 1A - Cell Col**). These bubble pockets existed in both cellular and acellular collagen. This was also seen in the migration/scratch assay. When the wells were scratched with the 1000 μL pipette tip, some of the cells were lifted from the collagen, and there was additional movement and folds added to the collagen.

The scratch-migration assay demonstrated that the most successful wound healing occurs when the cellular state of the top and bottom collagen agreed. The highest growth rate for all conditions was seen in the well with cellular collagen gels on both the top and bottom of the MCF-7 cells. However, the cellular state of the collagen only benefited the wound healing progression when both of the gels were cellular. When one collagen gel was cellular and one was acellular, this resulted in worse wound healing than the condition which had matching acellular collagen gels on the top and bottom of the scratch. This data shows that it is important to prioritize matching the collagen gel states to generate the most successful wound healing. Because cellular collagen is able to be used as a representative model of human skin, this makes the well conditions with cellular collagen on the bottom-most relevant to human application.

When cellular collagen was added to the top of scratch wounds in MCF-7 cells, the growth of the cells was too confluent in the scratch fields to count accurately. This result suggests that the addition of the cellular collagen layer on top of the wound rendered the scratch completely healed after 24 hours from a visual assessment standpoint. Additionally, it is important to recognize that even though when this condition had acellular collagen gels added to the top of the scratch wound, the healing was not as significant when compared to the cellular condition; however, the cell growth with the acellular gel was higher than the condition in which no collagen was added to the top of the scratch. Based on this, an acellular gel for wound healing may be more beneficial than not having one at all in situations in which a cellular collagen gel is not available. Overall, these results support the hypothesis that cell migration increases when exposed to cellular collagen, indicating a potential for applications of fibroblast-based wound healing mechanisms.

CONCLUSIONS

The cell attachment and viability assays showed that cellular collagen gels that are seeded with NIH/3T3 cells are able to promote higher levels of growth of cells than acellular collagen. To add to this finding, both cellular and acellular collagen gels resulted in more successful wound healing when collagen was added to the top of the scratch wound. Specifically, wound healing based on migration occurred most successfully for the cellular collagen conditions, but it was also noted that matching cellular states of the gels was important as well. This data supports our hypothesis that cellular collagen will promote wound healing because of its ability to model human epithelial cells and wound healing mechanisms. Our work provides a novel approach to explore the applications of collagen gels in wound healing. By creating a nontoxic model that placed collagen on both top and bottom of a scratch, we were able to more closely resemble a model of human epithelial tissue. This work is significant as it provides data on

the benefits of using cellular collagen as a model for epithelial tissue and as a method for wound healing, as well as developing a novel method of testing layered collagen gels for cell growth. Future directions to continue these methods would be to expand the number of biological and experimental replicates to increase the statistical resolving power, vary cellular concentrations within this experiment, and to conduct *in vivo* experiments that incorporate live animal models to determine if the results would continue to extend to live tissue.

ACKNOWLEDGMENTS

We thank the WPI Department of Biology and Biotechnology, as well as the EmpOwER grant awarded by the Women's Impact Network, for funding these research initiatives. We acknowledge and thank the Cell Engineering Research Equipment Suite (CERES) for allowing us to utilize their imaging and plate reading equipment to examine our experiments at a higher level. We also would like to thank Dr. Catherine Whittington and Dr. Jeanine Coburn from the WPI Department of Biomedical Engineering for consultation regarding materials and model construction.

REFERENCES

1. Dick, M. K., Miao, J. H., & Limaiem, F. (2022) *Histology, Fibroblast*. StatsPearls. Retrieved January 22, 2023, from *https://www.ncbi.nlm.nih.gov/books/NBK541065/*
2. Xue, M., & Jackson, C. J. (2015) Extracellular Matrix Reorganization During Wound Healing and Its Impact on Abnormal Scarring. *Adv Wound Care, 4*(3), 119–136. *https://doi.org/10.1089/wound.2013.0485*
3. Iyer, K., Chen, Z., Ganapa, T., Wu, B. M., Tawil, B., & Linsley, C. S. (2018) Keratinocyte Migration in a Three-Dimensional In Vitro Wound Healing Model Co-Cultured with Fibroblasts. *Tissue Eng Regen Med, 15*(6), 721–733. *https://doi.org/10.1007/s13770-018-0145-7*
4. Comșa, S., Cîmpean, A. M., & Raica, M. (2015) The Story of MCF-7 Breast Cancer Cell Line: 40 years of Experience in Research. *Anticancer Res, 35*, 3147–3154.
5. Foster, D. S., Jones, R. E., Ransom, R. C., Longaker, M. T., & Norton, J. A. (2018) The evolving relationship of wound healing and tumor stroma. *JCI insight, 3*(18). *https://doi.org/10.1172/jci.insight.99911*
6. Campbell, J. J., Husmann, A., Hume, R. D., Watson, C. J., & Cameron, R. E. (2017) Development of three-dimensional collagen scaffolds with controlled architecture for cell migration studies using breast cancer cell lines. *Biomaterials. 114*, 34–43. *https://doi.org/10.1016/j.biomaterials.2016.10.048*
7. Leibiger, C., Kosyakova, N., Mkrtchyan, H., Glei, M., Trifonov, V., & Liehr, T. (2013) First Molecular Cytogenetic High Resolution Characterization of the NIH 3T3 Cell Line by Murine Multicolor Banding. *J Histochem Cytochem, 61*(4), 306–312. https://doi.org/10.1369/0022155413476868
8. Vantangoli, M. M., Madnick, S. J., Huse, S. M., Weston, P., & Boekelheide, K. (2015) MCF-7 Human Breast Cancer Cells Form Differentiated Microtissues in Scaffold-Free Hydrogels. *PLoS ONE, 10*(8): *e0135426*. *https://doi.org/10.1371/journal.pone.0135426*
9. Meyer, M. (2019) Processing of collagen based biomaterials and the resulting materials properties. *Biomed Eng Online, 18*(1), 24-24. *https://doi.org/10.1186/s12938-019-0647-0*
10. Ramshaw, J. A. M. (2016) Biomedical Applications of collagens. *J Biomed Mater Res Part B, 104*(4), 665–675. *https://doi.org/10.1002/jbm.b.33541*
11. Ravi, M., Paramesh, V., Kaviya, S. R., Anuradha, E., & Solomon, F. D. P. (2015) 3D Cell Culture Systems: Advantages and Applications. *J Cell Physiol, 230*(1), 16–26. *https://doi.org/10.1002/jcp.24683*
12. Correa, S., Grosskopf, A. K., Lopez Hernandez, H., Chan, D., Yu, A. C., Stapleton, L. M., & Appel, E. A. (2021) Translational Applications of Hydrogels. *Chem Rev, 121*(18), 11385–11457. *https://doi.org/10.1021/acs.chemrev.0c01177*
13. Artym, V. V., & Matsumoto, K. (2010) Imaging Cells in Three-Dimensional Collagen Matrix. *Curr Protoc Cell Biol, 48*(1). *https://doi.org/10.1002/0471143030.cb1018s48*
14. Sotelo Leon, D., Williams, T., Wang, Z., Leyden, J., Franklin, A., Kaizawa, Y., Chang, J., & Fox, P. M. (2020) Analysis of Cell-seeded, Collagen-rich Hydrogel for Wound Healing. *Plast Reconstr Surg Glob Open, 8*(8), e3049–e3049. *https://doi.org/10.1097/gox.0000000000003049*
15. Liu, T., Qiu, C., Lu, H., Li, H., Zhu, S., & Ma, L. (2023) A novel recombinant human collagen hydrogel as minced split-thickness skin graft overlay to promote full-thickness skin defect reconstruction. *Burns, 49*(1), 169–181. *https://doi.org/10.1016/j.burns.2022.02.015*

ABOUT STUDENT AUTHORS

Claire Behning graduated from Worcester Polytechnic Institute in the Spring of 2023 with a Bachelor of Science in Biology and Biotechnology and a Bachelor of Science in Psychology. She enjoyed being able to expand her laboratory skills in cell culture while also being able to engage in the research process. As of 2023, she is applying to jobs as a high school biology teacher in public school districts, with the goal of applying her love of science and education to develop the next generation of scientists and lifelong learners.

Lia Kelly graduated from Worcester Polytechnic Institute in the Spring of 2023 with a Bachelor of Science in Biology and Biotechnology. She appreciated developing her cell culture skills and learning how to utilize biomaterials while experiencing the research process. As of 2024, she will be attending Tufts Cummings School of Veterinary Medicine in the fall as a DVM candidate with the hopes of continuing her involvement in research to improve human and animal health and medicine.

Emma Smith graduated from Worcester Polytechnic Institute in the Spring of 2023 with a Bachelor of Science in Biomedical Engineering. They enjoyed applying knowledge from their materials courses in a setting focused on biologics. As of 2023, they will be starting their MS in Biomedical Engineering at Worcester Polytechnic Institute in the Fall Semester. They plan to complete a research-based project to further their knowledge in functionalizing biomaterials, allowing them to utilize their knowledge to improve our understanding of disease and disease treatment.

PRESS SUMMARY

Fibroblasts are a cell type that helps to form connective tissue, which acts as the glue to support organs in the body and play a role in wound healing. Similarly, 3D collagen hydrogels have been studied as wound-healing systems for skin cells. In this study, MCF-7 skin cells were exposed to 3D collagen embedded with NIH/3T3 fibroblasts to explore the growth and attachment response as a representation of cell repair. The results showed that the most wound-healing effects were seen when MCF-7 cells were covered by and adhered to cellular 3D collagen in a "sandwich-like" configuration or if collagen was embedded with fibroblast cells. These methods can be continued *in vivo* to determine if the results would extend to live tissue.

Print Edition ISSN 1536-4585
Online Edition ISSN 2375-8732